

IRIS

Spécial

Édité par
l'Observatoire européen
de l'audiovisuel



OBSERVATOIRE EUROPÉEN DE L'AUDIOVISUEL
EUROPEAN AUDIOVISUAL OBSERVATORY
EUROPÄISCHE AUDIOVISUELLE INFORMATIONSSTELLE

La recherche de contenus audiovisuels



COUNCIL OF EUROPE
CONSEIL DE L'EUROPE

IRIS Spécial : La recherche de contenus audiovisuels

Observatoire européen de l'audiovisuel, Strasbourg 2008
ISBN : 978-92-871-6561-9
EUR 89

Directeur de la publication :

Wolfgang Closs, Directeur exécutif de l'Observatoire européen de l'audiovisuel
E-mail : wolfgang.closs@coe.int

Éditrice et coordonnatrice :

Susanne Nikoltchev, LL.M. (Florence/Italie, Ann Arbor/MI)
Responsable du département Informations juridiques
E-mail : susanne.nikoltchev@coe.int

Institution partenaire ayant collaboré à l'ouvrage :



Institut du droit de l'information (IViR)

Rokin 84
NL-1012 KX Amsterdam
Tél. : +31 (0) 20 525 34 06
Fax : +31 (0) 20 525 30 33
E-mail : website@ivir.nl
URL : <http://www.ivir.nl>

Assistante éditoriale :

Michelle Ganter

Marketing :

Markus Booms
E-mail : markus.booms@coe.int

Traduction/Relecture:

Caroline Bletterer, Alison Hindhaugh, Amélie Lépinard, Bernard Ludewig, Stefan Pooth, Erwin Rohwer, Roland Schmid, Dorothee Seifert-Willer, Nathalie Sturlèse, Candelaria Van Strien-Reney, Anne-Lise Weidmann

Photocomposition/Impression :

Pointillés, Hoenheim (France)

Éditeur :

Observatoire européen de l'audiovisuel
76 Allée de la Robertsau
F-67000 Strasbourg
Tél. : +33 (0)3 88 14 44 00
Fax : +33 (0)3 88 14 44 19
E-mail : obs@obs.coe.int
<http://www.obs.coe.int>

Veillez citer cette publication comme suit :

**Susanne Nikoltchev Ed., IRIS Spécial : La recherche de contenus audiovisuels
(Strasbourg, Observatoire européen de l'audiovisuel, 2008)**

© Observatoire européen de l'audiovisuel, 2008.

Chacune des opinions exprimées dans la publication est personnelle et ne peut en aucun cas être considérée comme représentative du point de vue de l'Observatoire, de ses membres ou du Conseil de l'Europe.

IRIS Spécial

La recherche de contenus audiovisuels

En 1989/90, Tim Berners-Lee a mis au point le World Wide Web (WWW) pour l'Organisation européenne pour la recherche nucléaire (CERN). Vingt ans plus tard, cette invention a déjà un impact sans précédent sur la façon dont nous traitons l'information. Le WWW est devenu une ressource stratégique de la vie privée comme de la vie professionnelle. Il offre à chacun une quantité et une diversité incroyables de contenus et nous permet également de l'utiliser pour partager nos propres créations avec autrui. Le WWW répond à notre attente concernant les informations qu'il fournit, il influe sur nos habitudes relatives au mode et au lieu de recherche des contenus, et il suscite des craintes quant à la question de savoir qui, de l'homme ou de la machine, domine l'autre.

Non seulement le CERN a donné naissance au Web, mais il lui a peut-être servi de modèle pour créer des quantités astronomiques d'informations. Le CERN devrait produire chaque année plus de 15 millions de giga-octets de données dans le cadre du traitement des centaines de millions de particules subatomiques dans le Grand Collisionneur de Hadrons (LHC), le plus puissant accélérateur de particules au monde, ce qui n'est pas sans ressembler à la masse sans cesse croissante d'informations circulant sur le WWW. Le moteur de ces deux systèmes se trouve dans notre volonté de collecter des informations.

Lors du lancement de la récente expérience du CERN visant à reconstituer le big bang originel, certains critiques ont rappelé que tout ce déploiement technologique n'allait pas sans risque. Certes, les responsables et les chercheurs scientifiques du CERN mettent tout en œuvre pour sécuriser l'expérience, néanmoins ils reconnaissent aussi qu'un seul esprit n'est plus en mesure de contrôler ou de superviser une expérience de cette ampleur. Par conséquent, la supervision est assurée par une vaste équipe, dont la responsabilité de chaque membre est limitée à une petite partie du projet.

Assiste-t-on à une évolution similaire avec le WWW ? Sommes-nous toujours en mesure de le maîtriser ou avons-nous créé un monstre de l'information qui échappe au contrôle du simple être humain ? Sommes-nous en train d'être submergés par un déluge d'informations ? Toutes ces questions doivent nous amener à réfléchir sur le rôle, voire les rôles essentiels que joue l'informatique dans notre vie :

I. Le volume d'information traité par l'informatique est ahurissant. L'une des contributions de la présente publication indique que les informations créées, stockées et copiées électroniquement en 2008 représentent l'équivalent de trois millions de fois le volume d'informations contenues dans tous les livres jamais écrits jusqu'à présent, en précisant que ce chiffre devrait être multiplié par six dans les deux ans qui viennent. Il est clair qu'aujourd'hui, la recherche d'un contenu audiovisuel pertinent est devenue un défi considérable. Le fait que nous parvenions encore à trouver ce que nous cherchons est dû avant tout aux outils électroniques tels que les moteurs de recherche et les guides électroniques de programmes.

Autrement dit, le foisonnement des informations électroniques nous contraint de développer de nouveaux outils au service de la recherche et de nous y fier.

II. Dans une situation où une prolifération de contenus issus de sources très diverses est stockée sur de multiples plateformes et mise à la disposition d'une population d'internautes de plus en plus nombreuse et diversifiée, il devient extrêmement difficile de fournir un système cohérent permettant à chacun d'obtenir des résultats satisfaisants. La recherche fructueuse de contenus audiovisuels est un défi d'une complexité accrue, car la plupart des outils de recherche sont basés sur le texte.

Toute approche cohérente visant à rendre les données adressables implique forcément une méthodologie cohérente d'identification et de classement des données. Il faut les classer et les étiqueter systématiquement, exactement comme c'était l'usage pour l'archivage des documents imprimés sur les rayons.

Autrement dit, le foisonnement des informations électroniques peut impliquer la nécessité de développer des systèmes universels d'identification des contenus.

III. La facilité avec laquelle on peut placer en ligne ce qu'on veut sur n'importe quel sujet a un impact significatif sur la possibilité de chacun de s'immiscer dans la vie d'autrui. Nous pouvons trouver en ligne les informations les plus extravagantes, voire les plus tendancieuses, mais aussi les plus pertinentes sur tout et sur tous. En même temps, nous pouvons nous-mêmes être la cible de ce type d'informations sans même en avoir connaissance.

Autrement dit, le foisonnement des informations électroniques rend impossible tout contrôle sur les informations à caractère personnel.

IV. La nécessité de canaliser et de filtrer le vaste flux des informations disponibles sur Internet pour que chacun puisse y trouver des éléments pertinents risque de se heurter au droit de la liberté d'expression. Lorsque les moteurs de recherche et les guides électroniques de programmes créent des listes de sources pertinentes, dans le même temps ils sélectionnent toujours les informations qui doivent être exclues de cette liste.

Autrement dit, l'utilisateur doit-il se satisfaire du fait que la sélection, au sein de ce foisonnement d'informations, soit faite par un tiers et à l'aide de systèmes automatisés ?

V. Considérant le rôle majeur des moteurs de recherche et des guides électroniques de programmes dans la recherche et la consultation des informations, n'y a-t-il pas nécessité de contrôler leur fonctionnement ? Qui intervient, le cas échéant, si une information se trouve exclue sans raison apparente par l'algorithme d'un moteur de recherche ? Selon quelles règles les moteurs de recherche et les guides électroniques de programmes produisent-ils leurs résultats ?

Autrement dit, le cadre réglementaire européen s'applique-t-il aux outils de recherche de contenus audiovisuels ?

VI. Ce n'est pas par hasard que le World Wide Web est aussi appelé « World Wild Web ». Il semble que les régulateurs se laissent inéluctablement distancer dans leurs efforts pour suivre les développements techniques et qu'ils n'aient jamais la réactivité requise pour accompagner le WWW dans sa progression fulgurante.

Que peut-on, ou plutôt que *doit-on* faire pour s'assurer que la fourniture d'informations en ligne soit conforme au système de valeurs européen ? Comment garantir, par exemple, l'équilibre délicat entre la liberté d'information et le droit à la vie privée ? N'y a-t-il pas un impérieux besoin de réguler ? Les mécanismes d'autorégulation et de corégulation pourraient-ils instaurer ce que le dispositif de régulation classique peine à mettre en place ?

Autrement dit, qui peut délimiter, si délimitation il y a, le cadre juridique de la recherche de contenus audiovisuels dans un foisonnement d'informations électroniques ?

Ces six points sont développés dans les six chapitres de ce numéro d'IRIS Spécial en mettant en lumière le contexte juridique dans lequel ils s'inscrivent. Le thème extrêmement complexe de cette publication a également fait l'objet de débats approfondis lors d'une table ronde co-organisée par l'Institut du droit de l'information de l'Université d'Amsterdam (IViR) et l'Observatoire européen de l'audiovisuel en avril 2008. L'essentiel de la discussion est présenté dans la synthèse des débats qui ouvre cet IRIS Spécial.

Nous souhaitons remercier tous les participants à cette table ronde pour leur contribution pertinente à la discussion. La liste des participants est fournie en fin de publication. Nous adressons tous nos remerciements à nos collègues de l'IViR pour leur aide précieuse dans l'organisation et la tenue de cette initiative, ainsi qu'aux auteurs des six articles et de la synthèse des débats. Nico van Eijk et Christina Angelopoulos, d'IViR, nous ont également apportés leur aide en pré-éditant les textes originaux anglais. Par ailleurs, nous rappelons que le succès de cette série IRIS Spécial est également le fait des traducteurs et relecteurs compétents et dévoués. Nous les remercions de l'excellent travail qu'ils ont fourni pour ce numéro d'IRIS spécial.

Strasbourg, novembre 2008

Wolfgang Closs
Directeur exécutif

Susanne Nikoltchev
Responsable du Département Informations juridiques

IRIS Spécial

**La recherche
de contenus audiovisuels**

SOMMAIRE

Table ronde sur la recherche audiovisuelle Synthèse des débats	1
<i>Christina Angelopoulos, Joris van Hoboken</i>	
Les défis technico-économiques liés aux moteurs de recherche audiovisuels	13
<i>Ramón Compañó</i>	
L'avenir de la navigation dans les contenus	33
<i>Janet Greco</i>	
Protection de la vie privée et quête du moteur de recherche audiovisuel parfait	39
<i>Michael Zimmer</i>	
L'incidence de la liberté d'expression sur la gouvernance des outils de recherche	53
<i>Joris van Hoboken</i>	
À la recherche des moteurs de recherche audiovisuels dans le cadre réglementaire de l'UE	71
<i>Peggy Valcke</i>	
Réguler les moteurs de recherche ? Le recours à l'autorégulation et à la corégulation dans le domaine de la recherche sur Internet	87
<i>Wolfgang Schulz</i>	

Table ronde sur la recherche audiovisuelle

Synthèse des débats

Christina Angelopoulos, Joris van Hoboken
Institut du droit de l'information (IViR), Université d'Amsterdam

Introduction

Le 12 avril 2008, l'Institut du droit de l'information de l'Université d'Amsterdam (IViR) et l'Observatoire européen de l'audiovisuel ont organisé une table ronde d'experts à Amsterdam sur le thème « La recherche audiovisuelle. Les défis juridiques qui résultent de l'abondance de contenu audiovisuel ». Cette initiative avait pour objectif de susciter un échange de points de vue sur l'avenir de la recherche audiovisuelle et les questions réglementaires connexes. Dans ce cadre, huit rapports distincts ont été présentés, chacun étant suivi d'un débat. La première série de rapports a dressé l'inventaire des défis concrets auxquels est actuellement confrontée la navigation dans les contenus. La seconde série était consacrée à l'analyse du cadre normatif et réglementaire au sein duquel opère ce secteur. Ce compte-rendu fournit une synthèse thématique des opinions exprimées et des conclusions auxquelles sont parvenus les participants.

Les rapports et les débats consécutifs sur les services de recherche audiovisuels, le marché et les lois régissant la chaîne de valeur des contenus audiovisuels ont mis en évidence le caractère extrêmement dynamique d'un environnement qui se développe très rapidement. L'un des principaux défis, pour l'avenir, sera de maintenir un cadre juridique et réglementaire régulièrement actualisé, tout en créant des conditions favorables au marché en pleine évolution des technologies et des services de recherche audiovisuels. Les questions fondamentales concernant la circulation libre et effective de l'information entre les fournisseurs de contenus et les usagers, ainsi que la question cruciale de la protection des données personnelles, devraient permettre d'alimenter le débat sur la nécessité éventuelle de réguler les moteurs de recherche audiovisuels. Étant donné que les services de recherche prennent une place de plus en plus centrale dans l'environnement de l'information, les législations et politiques futures devront prendre en compte ces nouvelles entités de façon plus appropriée.

1. Aspects concrets de la recherche audiovisuelle actuelle

1.1. Modèles commerciaux des moteurs de recherche et évolution technologique de la recherche

L'un des premiers éléments qui a retenu l'attention des participants concerne les tendances actuelles du secteur de la recherche audiovisuelle, tant au niveau des services de recherche en général que de l'exploitation du marché en particulier. La question des barrières bloquant l'accès des nouveaux acteurs sur le marché a notamment suscité une attention particulière. Les participants sont convenus que les principaux obstacles étaient, d'une part, le coût d'exploitation de l'exploration

et de l'indexation et, d'autre part, la réponse aux requêtes, comme le suggère le rapporteur Ramón Compañó¹.

Partant de là, une théorie a été avancée selon laquelle, hormis le coût financier, d'autres facteurs, tels que l'exploitation d'informations détaillées pour mieux cibler le public, pourraient également constituer des obstacles supplémentaires.

Les moteurs de recherche établis sont en train de compiler ce qu'on appelle des « bases de données de nos intentions² », qui leur permettent, à l'aide des profils constitués, d'associer les modèles d'une recherche individuelle à d'autres modèles, en vue d'optimiser l'efficacité et la popularité des produits. Cela leur permet également d'offrir une méthode efficace pour cibler les différents publics au profit des annonceurs publicitaires. Si cette pratique devient un critère indispensable pour pouvoir proposer un outil de recherche performant, les nouveaux venus qui n'ont pas eu la possibilité de collecter ce type de données seront désavantagés. Par conséquent, les participants ont évoqué le risque que les services personnalisés entraînent un effet de capture : actuellement, pour changer de prestataire de services, il suffit d'un clic. Mais si un moteur de recherche possède des informations personnelles sur ses utilisateurs, cela risque de les ancrer chez ce prestataire sur le long terme et desservir les concurrents qui souhaiteraient proposer des services similaires.

Cela signifie-t-il que le marché de la recherche sur Internet est un marché de monopole naturel ? Les participants étaient enclins à rejeter cette idée, notamment au vu des éléments suivants :

1. Google, actuellement n° 1 sur le marché, était en fait un nouveau venu. Les premiers moteurs de recherche tels qu'Alta Vista ont été, en définitive, évincés du marché.
2. Si Google est leader sur le marché américain et européen, ce n'est pas le cas dans d'autres régions du monde. Il est possible que la langue soit un obstacle à son expansion sur certains marchés régionaux.
3. Enfin, des moteurs de recherche thématiques font actuellement leur apparition, ce qui indique qu'il existe des créneaux sur le marché capables d'accueillir de nouveaux acteurs.

Le rapport de Thomas Roukens présentait le modèle commercial de Telenet, principal câblo-opérateur belge, fournissant ainsi une base concrète d'analyse des pratiques commerciales existantes.

La discussion a démarré sur le thème de la politique d'intégration verticale pratiquée par Telenet. Même si certains participants étaient critiques vis-à-vis de ce modèle d'intégration, ils ont reconnu que, du fait de la délimitation linguistique du marché flamand sur lequel opère Telenet, son cas est exemplaire sur ce point. Les participants ont relevé qu'au début, les radiodiffuseurs locaux et les producteurs de contenus étaient frileux par rapport à ce type de services numériques, persuadés qu'Internet allait balayer leurs usagers et leurs revenus. Telenet a tenté de surmonter cette réaction en créant un modèle collaboratif avec les fournisseurs de contenus et les radiodiffuseurs et en mettant l'accent sur la nature complémentaire des services supplémentaires qu'il fournissait.

Pour finir, non seulement il y a eu une augmentation des transactions pour iDTV (Integrated Digital Television), le service de vidéo à la demande (VoD - Video on Demand) de Telenet, mais les revenus des radiodiffuseurs locaux et des fournisseurs de contenus ont également progressé et leur position s'est consolidée. Les participants ont établi que dans une communauté linguistique restreinte, l'intégration verticale via la collaboration peut être l'unique moyen de survivre à la concurrence des grands radiodiffuseurs.

La discussion a également abordé la question du concept interactif évolutif des Guides de programmes électroniques (EPG - Electronic Programme Guides). Avant de lancer son propre service, Telenet était sceptique et se demandait si le consommateur moderne ne serait pas davantage attiré par quelque chose de plus clinquant et sophistiqué. Or, il semble que ce soit justement la simplicité du service proposé qui ait attiré les consommateurs. En outre, les utilisateurs ont largement fait usage des fonctionnalités leur permettant d'enregistrer leurs programmes. Finalement, l'EPG est devenu le pivot central de la plateforme interactive de Telenet. Les participants étaient perplexes quant à cette évolution, se demandant si, en fait, cela ne transformait pas l'EPG en système d'enregistrement.

Or, Telenet n'a pas cette impression, puisque les informations annexes sur les programmes ne sont pas simplement restées disponibles par le biais des EPG, mais sont massivement utilisées.

1) Voir l'article de Ramón Compañó dans cette publication.

2) Voir l'article de Ramón Compañó dans cette publication.

1.2 Le rôle des métadonnées

Le rapport de Janet Greco sur la nécessité d'avoir une source d'approvisionnement cohérente de métadonnées pour les EPG a orienté le débat sur les droits pertinents en matière de propriété intellectuelle. Les participants ont tenté de définir dans quelle mesure les droits de propriété intellectuelle existent sur les métadonnées et qui les détient. L'affaire *Magill*³ a été rapidement mentionnée. Cette affaire a confirmé la mise en place d'une licence obligatoire pour les chaînes de télévision en compensation de l'exercice de leurs droits exclusifs dans le cadre du droit d'auteur national, qui empêche les éditeurs de guides hebdomadaires de copier leurs listings. Cette affaire a fait apparaître qu'il existait effectivement des droits d'auteur sur les métadonnées sous forme de listings télévisés mais, en même temps, elle a également établi que, dans des circonstances exceptionnelles, l'exercice d'un droit exclusif de la part d'un titulaire pouvait constituer un acte abusif. Comme l'a fait remarquer un participant, ce raisonnement permet de conclure que le droit de la concurrence peut, dans certains cas, prévaloir sur la propriété intellectuelle.

Cette question a donné lieu à un échange d'arguments : d'une part, certains pensaient qu'il y avait de fortes incitations à protéger les droits de propriété intellectuelle sur, par exemple, les descriptifs et les classements des contenus audiovisuels. La référence d'un programme est une valeur ajoutée, car elle renforce la capacité de la société à comparer et à choisir. L'agent de cette plus-value devrait donc être en mesure de revendiquer un contrôle sur l'usage qui en est fait. Par conséquent, une législation telle que la directive UE sur les bases de données qui, en substance, protège les efforts investis dans la compilation de ces informations, est justifiée. D'autre part, une partie de l'assemblée affirmait que, d'un point de vue commercial, cette démarche crée des situations complexes. Un participant a exprimé l'idée selon laquelle on ne devrait pas reconnaître un droit d'auteur sur des informations telles que le titre d'un film ou le nom du metteur en scène, contrairement, par exemple, au synopsis, car ils ne sont que de simples questions de faits. Dans un environnement où le droit d'auteur est reconnu, un agrégateur qui met tout en œuvre pour regrouper des métadonnées cohérentes et correctes doit assurer la gestion des relations commerciales avec des sources de publications multiples et variées. En ce qui nous concerne, au niveau de la propriété des contenus, a-t-il poursuivi, on passe à côté de l'essentiel : l'objectif qui est de permettre aux consommateurs de trouver les contenus correspondant à leurs intérêts et munis de toutes les étiquettes requises d'un point de vue réglementaire.

En revanche, d'autres participants considèrent que le concept même d'une métadonnée cohérente est inaccessible dans un monde interconnecté. Il est difficile d'organiser les programmeurs ou les distributeurs de programmes, au sens où ils permettent la saisie et la classification systématiques d'informations selon les normes internationales. En outre, les contenus des bases de données concernées évoluent constamment et le taux de changement est en hausse. Pourtant, l'émergence de l'Internet devrait permettre de résoudre ces problèmes. Les moteurs de recherche sur Internet fonctionnent sur une base totalement différente, mais ils sont, sans aucun doute, mieux équipés pour affronter l'état de confusion dans lequel se trouvent les métadonnées. Grâce à l'usage des algorithmes, la recherche d'informations est efficace, même si elle est complexe. De légères incohérences telles que l'emploi ou non du deuxième prénom d'un acteur ne compromettent plus les résultats de façon significative. Comme l'expliquait un participant, les résultats des recherches ne doivent pas être exacts à 100 % pour être utilisables. Il leur suffit d'être suffisamment bons pour permettre l'utilisation d'un solide outil de recherche. Dans ce contexte, la question a été posée de savoir si les EPG n'étaient pas l'artefact d'un monde pré-Internet.

Enfin, le rapport de Ralph Traphöner qui analysait les mécanismes du projet THESEUS⁴, a démontré que la capacité des nouvelles technologies de recherche à traiter des blocs de données non cohérentes et à extraire des informations, ainsi que des métadonnées, à partir d'un environnement décentralisé et non contrôlé constituait l'une des innovations clefs de la technologie actuelle de recherche d'information. Par ailleurs, d'un point de vue technologique, les données audiovisuelles ne sont pas

3) Affaires conjointes C-241/91 P et C-242/91 P *RTE et ITP c. Commission (Magill)* [1995] ECR I-743.

4) Le projet THESEUS est un programme de recherche lancé par le ministère fédéral allemand de l'Économie et de la Technologie dans le but de développer une nouvelle infrastructure sur Internet permettant d'optimiser l'utilisation et l'exploitation des savoirs disponibles sur Internet. Actuellement, une trentaine d'institutions, d'universités et d'entreprises participent à ce programme sous la coordination d'empolis GmbH.

fondamentalement différentes des textes : pour obtenir des métadonnées qui décrivent un texte, on extrait des mots-clés des séquences. Dans le cas du matériel audiovisuel, le paradigme est identique, même si cela requiert des algorithmes plus sophistiqués et une capacité de calcul supérieure. L'inadéquation des EPG traditionnels ira croissante si, comme le prédit l'un des participants, il devient possible de rechercher non seulement des images (animées) par le biais de métadonnées, mais aussi par la comparaison des images elles-mêmes.

Cette dernière remarque a orienté la discussion sur le rôle que les consommateurs eux-mêmes pouvaient jouer au niveau du classement et de l'évaluation des contenus audiovisuels. Les exemples de Compact Disk Database (CDDB)⁵ et d'Internet Movie Database (IMDb)⁶ ont été cités. Dans ce type de modèle, la collecte des données est placée en partie dans les mains des utilisateurs, qui sont fortement incités à fournir des informations et sont les mieux placés pour connaître leurs besoins. De tels systèmes pourraient donc s'avérer hautement opérationnels. Mais là aussi, la question de la propriété refait surface : il est révélateur que CDDB ait été finalement absorbé, vendu et relancé sous le nom de Gracenote, une filiale à 100 % de Sony Corporation of America, tandis qu'IMDb a été racheté par Amazon.

Quoiqu'il en soit, les participants ont suggéré que la solution ne pouvait venir que du marché. Les sociétés de bases de données vont probablement s'imposer et devenir les gardiennes de l'accès aux métadonnées. L'un des participants a même envisagé la possibilité que la solution soit trouvée de façon concurrente : contrairement au système paneuropéen unique de classification des jeux vidéo PEGI, l'Europe possède 27 différents systèmes de classement dans le domaine du cinéma. C'est probablement un avantage qui donne aux consommateurs une vision plus objective, plutôt que de l'obliger à accepter un seul point de vue comme une vérité immuable.

2. Perspectives en matière de droits fondamentaux dans la recherche audiovisuelle

2.1. Contenus audiovisuels en ligne et vie privée

Le rapport de Michael Zimmer sur le marché faustien auquel nous sommes actuellement confrontés, puisque la technologie de recherche audiovisuelle élargit l'accès et la consultation des informations tout en menaçant la vie privée des utilisateurs⁷, a été suivi d'un débat animé. Au cours de cette discussion, l'assemblée s'est interrogée sur la connaissance qu'a un utilisateur moyen des menaces qui pèsent sur sa vie privée et sur son intérêt à se protéger lui-même. Une objection a néanmoins été émise, à cet égard, en rappelant que les atteintes à la vie privée pouvaient s'avérer difficiles à percevoir par des particuliers à un niveau abstrait et décontextualisé. En revanche, si le problème touche l'utilisateur de façon personnelle, c'est-à-dire si sa propre perception de la vie privée est compromise dans la réalité, il comprendra beaucoup mieux les problèmes liés au respect de la vie privée. En définitive, les moteurs de recherche sont tellement intégrés à notre vie quotidienne que la sensibilisation des usagers sur les menaces qu'ils représentent pour la vie privée s'est émoussée par le fait d'une surexposition. La confiance que nous inspirent les moteurs de recherche est la clef de leur succès.

Par ailleurs, l'hypothèse selon laquelle ce problème serait aggravé dans le cas de ce qu'on appelle les « exclus de l'information » a été émise. Il est à craindre que ceux qui ne sont pas connectés et, par conséquent, ne profitent pas des avantages de la technologie, ne se retrouvent, en outre, en position de faiblesse quant au respect de leur vie privée sur le Net. La procédure de notification et de retrait adoptée par les autorités compétentes de la plupart des pays atteint vite ses limites avec ceux qui ne possèdent pas ou ne maîtrisent pas les moyens technologiques permettant de s'informer sur les risques d'intrusion dans la vie privée.

5) Compact Disc Database (CDDB) est une base de données accessible sur Internet permettant à des applications logicielles de consulter des informations sur des disques audio compacts. Édité sous licence GNU General Public, le logiciel d'origine derrière CDDB était basé sur les contributions volontaires des utilisateurs. Le projet a finalement été incorporé en 1998 en tant que CDDB LCC. Il a ensuite été racheté par Escient, un constructeur d'électronique multimédia high-tech, et rebaptisé Gracenote.

6) Internet Movie Database (IMDb) est une base de données en ligne d'informations et de photos concernant le cinéma, les émissions de télévision, les acteurs, le personnel de production et les jeux vidéo. IMDb a vu le jour le 17 octobre 1990 et a été racheté par Amazon.com en 1998.

7) Voir l'article de Michael Zimmer dans cette publication.

Enfin, l'efficacité de la tendance actuelle, qui consiste à éduquer les utilisateurs comme moyen de défense, a été remise en cause. Certains ont fait remarquer que cette démarche présupait non seulement qu'un utilisateur était conscient de l'existence du problème, mais qu'il était aussi capable d'assimiler une foule d'informations extrêmement pointues sur une multitude de sujets, depuis le droit d'auteur et la législation sur le respect de la vie privée jusqu'à la protection des consommateurs.

En ce qui concerne les solutions possibles, certains participants ont évoqué l'obligation d'une intervention des milieux universitaires et des pouvoirs publics. Les représentants des pouvoirs publics devraient pointer les problèmes et défendre les droits concernant toute la communauté d'utilisateurs, sans proposer à quiconque en particulier de mobile personnel pour une action immédiate. Leur rôle pourrait être particulièrement positif auprès des membres les plus démunis de la société de l'information. D'autres mesures réglementaires ont également été envisagées, par exemple la mise en place de certaines obligations telles que le floutage des visages pour des services tels que Google Street View⁸ ; bien entendu, ce type de mesures prend le problème à revers en instaurant une protection par défaut. La proposition de mettre en place une information pédagogique de service public visant à favoriser l'édification personnelle des utilisateurs a également été avancée.

Les participants ont souligné que de nombreuses questions liées à la vie privée, qui se posent du fait de l'apparition de services de recherche audiovisuels de plus en plus sophistiqués, sont, d'une façon ou d'une autre, couvertes par le droit communautaire en matière de protection des données. À cet égard, l'avis sur les aspects de la protection des données liés aux moteurs de recherche émis récemment par le groupe de travail « Article 29 » sur la protection des données a été mentionné pour ses conclusions notables, en particulier en ce qui concerne l'utilisation par les moteurs de recherche des logiciels de reconnaissance faciale : « Les fournisseurs de moteurs de recherche spécialisés dans la création d'opérations à valeur ajoutée telles que les profils de personnes physiques (*people search engines*) et les logiciels de reconnaissance faciale sur images et contenus audiovisuels, doivent veiller à ce que leur traitement soit légitime, notamment par le consentement des personnes concernées, et doivent se conformer par ailleurs à toutes les exigences de la Directive sur la protection des données, telles que l'obligation de garantir la qualité des données et un traitement équitable »⁹.

Cela replace la protection à la charge du prestataire de service. À cet égard, une autre solution consisterait à impliquer les sociétés privées qui explorent le Web à la recherche de données compromettantes pour le respect de la vie privée. Néanmoins, les participants soulignent l'inconvénient de cette option qui ne profiterait qu'aux acteurs financièrement solides.

Par ailleurs, sur un plan plus technique, des solutions de contrôle des données personnelles et de leur traitement peuvent également être envisagées. Dans cette optique, il a été proposé un type de cryptage personnel et de système de gestion des droits numériques (*digital rights management - DRM*) permettant de contrôler les contenus téléchargés sans limiter la possibilité de partager les informations avec une communauté de contacts en ligne. D'ici là, on pourrait voir émerger l'application de licences lisibles par les ordinateurs et protégeant les données (de façon identique à la suite de licences Creative Commons, par exemple, qui est conçue pour protéger les contenus relevant du droit d'auteur). En tout état de cause, les participants ont pu identifier de multiples défis. La mise en œuvre constitue une pierre d'achoppement (comment contraindre les usagers à respecter les restrictions spécifiées ?), tandis que la mise en place du niveau requis de régularité semble également difficile (comment imaginer tous les contextes dans lesquels une image peut être utilisée afin de l'encoder dans un système d'information protecteur ?). À long terme, bien sûr, si la situation implique une contextualisation croissante de la vie privée, il faudra finir par établir des modèles permettant un certain degré d'automatisation.

Par ailleurs, il faudrait éviter une protection excessive de la vie privée et respecter le libre flux des informations, étant donné que les contenus en ligne constituent une part de notre sphère publique. En fait, au cours de la discussion, l'opinion a été avancée qu'aucun concept de protection de la vie privée n'était garanti : les moteurs de recherche permettent de pénétrer dans la vie privée d'autrui

8) The Guardian, *Google Blurs the Privacy Issue* (mai 2008), disponible sur :

<http://www.guardian.co.uk/business/2008/may/13/google.digitalmedia>, consulté le 15 juin 2008.

9) Groupe de travail « Article 29 », Avis sur les aspects de la protection des données liés aux moteurs de recherche, WP 148, 4 avril 2008.

mais, en même temps, les utilisateurs devraient être conscients du fait que les informations mises en ligne sont placées dans le domaine public. Une réglementation détaillée peut apparaître comme une réaction excessive, avec un bénéfice minime pour les droits effectifs liés à la vie privée, alors qu'un simple ajustement de la conduite pourrait fournir des résultats plus efficaces.

Enfin, il faut noter que le concept susmentionné de « base de données de nos intentions »¹⁰ a provoqué des inquiétudes en lien avec la vie privée. La réponse proposée affirmait que les compromis de ce type étaient des composants nécessaires de la société de l'information.

En fin de compte, il faut instaurer un dispositif permettant de maintenir l'équilibre entre la valeur liée à la fourniture d'informations efficaces et l'intrusion dans la vie privée que cela risque d'entraîner. Dans la recherche nomade, par exemple, les requêtes les plus fréquentes portent sur la météo, diverses informations locales et des cartes, qui sont manifestement en mesure de révéler le lieu où se trouve la personne, mais qui, en même temps, fournissent des informations utiles de façon instantanée. Par conséquent, il semble que notre conception de la vie privée devra évoluer, à l'avenir, pour s'adapter aux nouvelles possibilités offertes par la technologie.

En fait, la majorité des participants est convenue que la protection de la vie privée et, en particulier, des données personnelles, ne devrait pas se focaliser sur un mode de protection par blocage, mais envisager une approche plus souple. Il est difficile d'envisager un régime légal, voire technologique, global qui puisse régir toutes les différentes sortes d'interaction entre les gens, notamment pour les fonctions de recherche qui ignorent les barrières traditionnelles autorisant ce type d'interaction différenciée, c'est-à-dire qui suppriment « la vie privée par l'obscurité »¹¹.

2.2. Contenus audiovisuels en ligne et liberté d'expression

Au vu de ce qui précède, le besoin de clarifier la position juridique des moteurs de recherche en tant que menace pour la liberté d'expression devient extrêmement pressant. Le rapport de Joris van Hoboken a permis d'ouvrir le débat sur ce sujet.

Dans un premier temps, les participants ont soulevé certaines questions concernant la capacité du cadre juridique actuel à traiter les problèmes de liberté d'expression en ligne. Il est vrai que l'article 10 de la CEDH¹² a souvent été critiqué pour son caractère obsolète. Même si la Cour européenne des Droits de l'Homme s'est efforcée de maintenir une approche dynamique et évolutive, elle n'a pas toujours été en mesure de suivre l'évolution des mutations technologiques. En ce qui concerne la question de la protection de l'accès des fournisseurs d'information aux moteurs de recherche, les participants ont souligné que l'article 10 n'impliquait pas l'instauration d'un droit généralisé de toucher un public¹³. Néanmoins, cette approche peut indéniablement exclure certains fournisseurs d'information de l'accès équitable aux moyens de communication les plus efficaces, c'est-à-dire les plateformes de recherche. En définitive, l'accès à un public volontaire est un préalable à l'exercice de la liberté d'expression : s'il fait défaut, votre voix tombe dans le vide.

À titre d'alternative aux règles strictes en matière d'accès, on a avancé l'idée qu'il serait peut-être judicieux de considérer diverses sources de réglementations plus souples, telles que les recommandations du Conseil de l'Europe¹⁴. Enfin, les participants ont exprimé l'hypothèse que les devoirs et responsabilités assortis à la liberté d'expression en vertu de l'article 10 peuvent servir de base pour définir la responsabilité des moteurs de recherche, tout au moins dans une démarche de corégulation ou d'autorégulation.

10) Voir ci-dessus p. 1.

11) Voir l'article de Michael Zimmer dans cette publication.

12) Article 10 de la Convention de sauvegarde des Droits de l'Homme et des Libertés Fondamentales (Convention européenne des Droits de l'Homme dans sa version amendée) (signée le 4 juin 1950 et entrée en vigueur le 3 septembre 1953 ; ci-après : CEDH).

13) Voir *Stafford c. Royaume-Uni*, Arrêt de Grande Chambre de la Cour européenne des Droits de l'Homme du 28 mai 2002, paragraphe 68 ; *Appleby et Autres c. Royaume-Uni*, Arrêt de la Cour européenne des Droits de l'Homme (quatrième section) du 6 mai 2003, paragraphes 47 et 48 ; en outre, des questions similaires liées à la radiodiffusion apparaissent *inter alia* dans : *VgT Verein gegen Tierfabriken c. Suisse*, Arrêt de la Cour européenne des Droits de l'Homme (seconde section) du 28 juin 2001 et *Haider c. Autriche*, Décision d'irrecevabilité de la Cour européenne des Droits de l'Homme (première chambre) du 18 octobre 1995, requête n° 25060/94.

14) À titre d'exemple, voir l'article de Joris van Hoboken dans cette publication.

Ensuite, les participants se sont penchés sur la récente recommandation du Conseil de l'Europe concernant les filtres Internet¹⁵. Au début de la première ligne directrice, la règle spécifiant que « lorsqu'un utilisateur est confronté à un filtre, il doit être informé qu'un filtre est activé et, s'il y a lieu, il doit savoir reconnaître et contrôler le niveau de filtrage auquel est soumis le contenu qu'il consulte » a soulevé de vives réactions. L'un des participants a estimé que cette exigence était excessive, notamment si l'on considère l'usage généralisé, par exemple, des filtres anti-spam.

Bon nombre de ces filtres sont mis en place sans que les utilisateurs ne le sachent, ni ne l'acceptent. Et pourtant, la plupart d'entre eux estime que c'est un service utile. Il n'y a donc pas lieu de préconiser la nécessité d'informer l'utilisateur de la mise en place d'un tel filtre.

Enfin, du point de vue de l'utilisateur, il est clair que le droit d'accès à l'information acquiert un surcroît d'importance dans l'environnement en ligne. En fait, l'article 19 de la Déclaration universelle des Droits de l'Homme¹⁶ prévoit le droit de « chercher [...] les informations et les idées par quelque moyen d'expression que ce soit » ; il s'agit là d'un élément insuffisamment développé qui est souvent oublié. Manifestement, il est de l'intérêt de l'utilisateur de rechercher massivement parmi les informations qui sont disponibles en ligne. Mais il est important de comprendre que ce que garantit l'article 19, c'est-à-dire la liberté de rechercher des informations sans entrave, est foncièrement différent du droit effectif d'accès à l'information. Comme l'a fait remarquer l'un des participants, cette distinction prend toute son importance si l'on considère les principaux éléments de la discussion sur le respect de la vie privée mentionnés ci-dessus : lorsqu'une société a développé des outils de recherche assez sophistiqués pour leur permettre de consigner chacune des opérations effectuées par ses utilisateurs, le droit d'accès effectif peut émerger comme une arme puissante et redoutable. Cela fait apparaître une zone de tension entre la liberté d'information et la vie privée : chaque droit fournissant, dans le meilleur des cas, les mécanismes requis pour contrebalancer et juguler les excès de l'autre.

3. Aspects réglementaires de la recherche audiovisuelle

3.1. Place de la recherche audiovisuelle dans le droit et la réglementation actuels

Un certain nombre de propositions a été avancé concernant la meilleure approche pour traiter du régime réglementaire de la recherche et des outils de navigation. Le rapport de Peggy Valcke expliquait le traitement fragmentaire actuellement en vigueur : le cadre réglementaire actuel de l'Union européenne en matière de communications électroniques (Directive cadre sur les réseaux et services de communications électroniques, ci-après RSCE) n'est pas approprié pour réglementer les moteurs de recherche en général, ni les moteurs de recherche audiovisuels en particulier. L'article 2 de la Directive « cadre » spécifie formellement que le terme « services de communications électroniques » exclut les services de la société de l'information. Par conséquent, les moteurs de recherche ne sont pas couverts. De même, ils ne peuvent être considérés comme des « ressources associées », que l'article 2 e) de la Directive « cadre » définit comme « ressources associées à un réseau ou un service de communication électronique qui permettent ou facilitent la fourniture des services via ce réseau ou ce service¹⁷. » Les EPG, en revanche, sont spécifiquement mentionnés par la directive comme étant des ressources associées. En tout état de cause, la Directive cadre RSCE était destinée, du point de vue des aspects liés à la radiodiffusion, à réglementer les plateformes de télévision intégrées verticalement, où un opérateur contrôle toute une série de dispositifs. Dans le cas d'Internet, par contre, la chaîne est structurée différemment.

Les participants ont repéré un emplacement, dans la Directive « cadre », qui pourrait laisser une place aux moteurs de recherche, du moins en ce qui concerne la protection des consommateurs. Le texte proposé pour modifier l'article 20 (5) de la Directive sur les services universels¹⁸ tel qu'il est

15) Recommandation CM/Rec(2008)6 du Comité des Ministres aux Etats membres sur les mesures visant à promouvoir le respect de la liberté d'expression et d'information au regard des filtres Internet, disponible sur : [https://wcd.coe.int/ViewDoc.jsp?Ref=CM/Rec\(2008\)6](https://wcd.coe.int/ViewDoc.jsp?Ref=CM/Rec(2008)6), consulté le 1^{er} juillet 2008.

16) Déclaration universelle des Droits de l'Homme (adoptée le 10 décembre 1948) Rés. AGNU 217 A (III), art. 19.

17) Directive 2002/21/CE du Parlement européen et du Conseil du 7 mars 2002 relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques (Directive « cadre ») [2002] JO L 108/33.

18) Directive 2002/22/CE du Parlement européen et du Conseil du 7 mars 2002 concernant le service universel et les droits des utilisateurs au regard des réseaux et services de communications électroniques (Directive « service universel ») [2002] JO L108/51.

formulé est le suivant : « Les États membres veillent à ce que, lorsqu'un contrat est conclu entre un abonné et une entreprise fournissant des services et/ou des réseaux de communications électroniques, l'abonné soit clairement informé, avant la conclusion du contrat et régulièrement par la suite, de toute restriction imposée par le fournisseur quant aux possibilités d'accéder à, et de distribuer, des contenus licites ou d'utiliser des applications et services licites de son choix. » La réforme en cours de la Directive RSCE pourrait offrir une opportunité de procéder à des changements législatifs en vue d'élargir la notion de ressources associées, telles qu'elles sont définies ci-dessus, de façon à y inclure les moteurs de recherche. En outre, l'article 20 (5) devrait être reformulé pour être également applicable dans les situations où il n'existe pas de contrat en tant que tel, mais où l'utilisateur utilise *de facto* un service. Ceci ouvrirait la voie pour obliger les moteurs de recherche à fournir davantage de transparence sur des questions telles que, par exemple, les méthodes de classement ou les liens sponsorisés. Cette suggestion a provoqué des réticences, surtout en regard du fait que le même effet pourrait être obtenu par l'autorégulation des moteurs de recherche.

Enfin, les participants ont également analysé dans quelle mesure les dispositions de la Directive sur le commerce électronique¹⁹ pouvaient avoir un impact sur les moteurs de recherche. Comme certains l'ont fait remarquer, alors que cette directive s'applique d'une façon générale aux moteurs de recherche, elle n'offre en réalité que très peu d'éléments en termes d'obligations effectives. Les articles 12 à 14 de la directive prévoient des règles « refuge » pour trois types d'intermédiaires : le stockage dit « *caching* », l'hébergement et le simple transport²⁰. L'article 21 de la directive spécifie clairement que ces dispositions ne couvrent pas « la responsabilité des fournisseurs de liens d'hypertexte et de services de moteur de recherche ».

Au contraire, cet article exclut les moteurs de recherche de la procédure de rapport bisannuel sur l'application de la directive, visant en particulier à analyser « la nécessité de présenter des propositions relatives à la responsabilité des fournisseurs de liens d'hypertexte et de services de moteur de recherche, les procédures de notification et de retrait (*notice and take down*) et l'imputation de la responsabilité après le retrait du contenu²¹. »

À cet égard, l'argument a été avancé selon lequel, nonobstant le refuge aménagé par l'article 13 de la Directive sur le commerce électronique pour le *caching*, dans certains cas, le *caching* des moteurs de recherche pourrait impliquer une responsabilité, notamment en regard de la pratique habituelle des moteurs de recherche de stocker des pages Web dans le but précis de pouvoir les proposer lorsque l'accès aux pages d'origine aura été supprimé. Sur ce point, des analogies ont été faites avec des dispositions similaires de la Directive sur le droit d'auteur, qui comporte également un article consacré au *caching*²². Dans ce contexte il a été fait référence à l'affaire *Copiepresse* qui se déroule actuellement devant les tribunaux belges. Certains ont fait remarquer que le Tribunal belge de Première instance²³ avait estimé que le *caching* pratiqué par Google s'apparentait à une reproduction et une diffusion publique sans autorisation et, de ce fait, constituait une infraction au droit d'auteur. Le tribunal a considéré que dans la mesure où le *caching* pratiqué par un moteur de recherche consistait en une opération de « stockage automatique, intermédiaire et temporaire de cette information, fait dans le seul but de rendre plus efficace la transmission ultérieure de l'information à la demande d'autres destinataires du service²⁴ », la règle refuge aménagée par la Directive sur le commerce électronique était applicable d'une façon générale.

En tout état de cause, les participants ont conclu que, dans ce cas particulier, l'enjeu n'était pas le stockage temporaire des pages « en cache » en tant qu'élément du processus d'indexation, mais la visibilité des pages en cache pour les utilisateurs, considérée comme une infraction au droit d'auteur.

19) Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« Directive sur le commerce électronique ») [2000] JO L178/1.

20) Articles 12 à 14 de la Directive sur le commerce électronique.

21) Article 21 de la Directive sur le commerce électronique.

22) Article 5 de la Directive 2001/29/CE du Parlement européen et du Conseil du 22 mai 2001 sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information [2001] JO L167/10.

23) *Google Inc. c. Copiepresse SCRL* (RB (Brussels)) Tribunal de Première Instance (Bruxelles) 13 février 2007, [2007] E.C.D.R. 5 2007 WL 1623283.

24) Article 13 de la Directive sur le commerce électronique.

Bien entendu, l'option la plus évidente pour une régulation des moteurs de recherche audiovisuels serait, le cas échéant, de l'intégrer à la Directive sur les services de médias audiovisuels (SMAV)²⁵. C'est, en effet, cette directive qui régit les questions culturelles et de contenus, contrairement à la Directive cadre RSCE, qui traite des questions de transmission et de concurrence. Or, le problème est que la Directive SMAV ne couvre pas, actuellement, les moteurs de recherche audiovisuels²⁶, qu'elle vient d'être révisée récemment, en décembre 2007, et qu'il est peu probable que la Commission ne la révisé à nouveau avant quelques temps. Néanmoins, ce sujet a donné lieu à un débat théorique, principalement axé sur l'interprétation du terme de « responsabilité éditoriale ».

La responsabilité éditoriale est une condition spécifique du service de média audiovisuel, tel qu'il est visé à l'article 1 de la Directive SMAV, qui a été définie comme étant le principal critère pour inclure les moteurs de recherche dans son champ d'application²⁷. Le même article définit la responsabilité éditoriale comme « l'exercice d'un contrôle effectif tant sur la sélection des programmes que sur leur organisation, soit sur une grille chronologique [...], soit sur un catalogue [...] ».

La discussion a permis d'établir que l'intention était, du moins de la part du législateur lors de la rédaction de cette directive, d'exclure les moteurs de recherche de cette définition. Il semble que les travaux préparatoires fassent apparaître clairement que l'objectif visé était les activités des radio-diffuseurs, c'est-à-dire la sélection des programmes et leur intégration dans une grille chronologique ou un catalogue. Cette activité est distincte de la création automatique d'une liste par le biais d'un algorithme. Le considérant 19 de la Directive SMAV spécifie que « la définition du fournisseur de services de médias devrait exclure les personnes physiques ou morales qui ne font que diffuser des programmes dont la responsabilité éditoriale incombe à des tiers. » Les « transporteurs », notamment les moteurs de recherche, ont donc été formellement exclus.

Cependant, certains participants ont soutenu que cette définition pouvait être suffisante pour inclure les moteurs de recherche. Selon cette thèse, l'utilisation d'un algorithme est, en fait, la preuve d'un contrôle effectif car, en définitive, un algorithme qui ne tiendrait pas compte de l'exclusion des incitations à la haine devrait être modifié en conséquence. Contrairement aux hébergeurs, qui ne sont pas tenus de se préoccuper du contenu des informations stockées sur leurs serveurs, dans la mesure où ceux-ci sont légaux²⁸, les algorithmes des moteurs de recherche se préoccupent de la nature et du contenu des informations présentées sur les sites Web dont ils fournissent les références, puisque leur mission consiste précisément à fournir aux utilisateurs des informations pertinentes en réponse à leurs requêtes.

Les détracteurs de ce raisonnement s'appuyaient sur la conception technologique du mode de fonctionnement effectif des moteurs de recherche. Ils ont fait observer que c'était le contenu qui contrôlait le moteur de recherche et non l'inverse. Par conséquent, un moteur de recherche est différent d'un catalogue ou d'un répertoire derrière lequel se trouve une équipe rédactionnelle humaine qui, incontestablement, exerce une responsabilité éditoriale au sens visé par l'article 1 de la Directive SMAV. De même, il convient de faire la distinction avec les plateformes de recherche verticales telles que, par exemple, Google News, car dans ces cas-là, c'est l'opérateur du moteur qui décide du type de contenu particulier que le moteur recherche. Ainsi, certains pensent que cela conforte la thèse selon laquelle des plateformes telles que YouTube exercent une responsabilité éditoriale. Une fois de plus, une perception analogique dans le domaine du droit d'auteur pourrait avoir de sérieuses implications pour la responsabilité des moteurs de recherche. L'affaire *Viacom*²⁹ a été citée à titre d'exemple : si l'on admet

25) Directive 2007/65/CE du Parlement européen et du Conseil du 11 décembre 2007 modifiant la Directive 89/552/CEE du Conseil visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à l'exercice d'activités de radiodiffusion télévisuelle [2007] JO L332/27.

26) Voir l'article de Peggy Valcke dans cette publication.

27) Voir l'article de Peggy Valcke dans cette publication.

28) Article 14 de la Directive sur le commerce électronique.

29) L'affaire *Viacom* (*Viacom International Inc., v. YouTube, Inc.*, n° 07 Civ. 2103 (S.D.N.Y., 13 mars 2007)) est une affaire en cours devant les tribunaux des États-Unis qui suscite un grand intérêt, dans laquelle le consortium des médias Viacom (« Video & Audio Communications ») réclame à Google et à sa filiale YouTube 1 milliard USD pour violation massive des droits d'auteurs. Viacom affirme que YouTube héberge sans autorisation de nombreux clips protégés, téléchargés par les utilisateurs à partir ses programmes de divertissement. D'autre part, YouTube revendique l'application de la disposition refuge du *Digital Millennium Copyright Act* (17 U.S.C. § 512(c)(1) (2000)) (Loi sur le droit d'auteur à l'ère du numérique), en vertu de laquelle les FAI ne sont pas tenus de surveiller leurs sites pour repérer d'éventuels contenus illicites, mais doivent retirer ces derniers sans délai dès lors qu'on leur a notifié leur existence. Viacom maintient que les conditions requises pour l'application de cette exception refuge ne sont pas réunies.

qu'un prestataire de service spécifique exerce un contrôle éditorial, cela l'exclut des dispositions refuges du *Digital Millennium Copyright Act* (loi sur le droit d'auteur à l'ère du numérique) ou des législations similaires en Europe, en le rendant responsable de la prévention d'éventuelles infractions au droit d'auteur du fait des contenus vers lesquels il crée des liens et du fait de la nature illégale ou pernicieuse desdits contenus.

Enfin, d'un point de vue commercial, la démarche de Telenet en matière de responsabilité éditoriale a également fait l'objet d'un débat. Les participants ont commencé par noter que, du fait de la vaste définition de la radiodiffusion dans la législation belge, celle-ci inclut également les services à la demande, ce qui signifie que Telenet ne saurait en aucun cas échapper à la responsabilité éditoriale. Dans ce cadre juridique, Roukens a également expliqué que Telenet avait tenté d'instaurer un équilibre entre le désir logique d'une société de limiter sa responsabilité et l'attente des clients. L'étude indique assez clairement que le public n'est pas familiarisé avec les services à la demande. Par conséquent, il attend le même type de services sur la plateforme IDTV que ceux qui sont fournis à la location d'un film. Telenet a répondu à cette attente en installant, par exemple, des systèmes dans le décodeur qui filtrent les films déconseillés pour les mineurs.

3.2. Autorégulation et corégulation des moteurs de recherche

Actuellement, en raison de l'absence d'une réglementation claire et centralisée de la part de l'État, les fournisseurs de moteurs de recherche doivent se fixer eux-mêmes des règles et des normes. Mais, comme le fait observer Wolfgang Schulz dans son rapport, la pratique d'autorégulation est loin d'être transparente et son efficacité est douteuse. L'incertitude juridique qui en découle peut être difficile à gérer, en particulier du point de vue des petits prestataires. D'une façon générale, cette pratique pourrait produire un effet paralysant³⁰. Au cours de cette table ronde, certains ont observé que les moteurs de recherche ont eux-mêmes exprimé la volonté d'une réglementation plus précise, qui leur fournirait une base plus solide pour résoudre des questions telles que, par exemple, savoir si la prévisualisation d'une photo dans une liste de résultats constitue ou non, en soi, une infraction au droit d'auteur.

Quoiqu'il en soit, dans la mesure où elle existe, la réglementation des moteurs de recherche implique principalement des procédures de notification et de retrait. Il est à noter que, pour le moment, aucune disposition de remise en place n'est en vigueur, ce qui pourrait s'avérer problématique. Néanmoins la discussion a fait apparaître une remise en cause des avantages de cette méthode de notification et de retrait comme défense contre les contenus litigieux sur Internet. En fin de compte, comme l'a noté un participant, si la cible se limite à la simple référence, qui est créée automatiquement en utilisant les informations provenant de tiers, tandis que le contenu original reste en ligne, notre action s'apparente, dans le meilleur des cas, au type même de l'entreprise inutile. Il serait certainement plus efficace de cibler le site Web d'origine. Dans ce contexte, la proposition a été faite d'aligner la démarche vis-à-vis des moteurs de recherche sur celle des services de *caching* plutôt que celle des services d'hébergement.

Hormis ces réflexions, une question a été posée concernant le code du *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter* (Association d'autorégulation volontaire des services multimédias, ci-après FSM)³¹. Ce dernier a mis en place un code de conduite d'autocontrôle que la plupart des grands moteurs de recherche ont adopté en Allemagne depuis 2004. Il semblait y avoir une certaine confusion parmi les participants au sujet du caractère volontaire de ce régime. Les moteurs de recherche eux-mêmes déclarent le contraire dans certaines situations, c'est-à-dire qu'ils prétendent que le régime d'autorégulation est imposé par l'État. Même si un certain nombre de sites Web réputés non conformes dans le cadre de ce régime restent, dans une large mesure, accessibles par le biais de certains moteurs participants, Google, pour sa part, annonce sur la page de résultats que les références des résultats retirés font suite à une injonction légale. Les participants en sont arrivés à la conclusion que la réponse résidait dans l'interprétation du terme « volontaire » : lorsque les prestataires filtrent, sur la base de

30) Voir l'article de Joris van Hoboken dans cette publication.

31) Code de conduite pour les moteurs de recherche allemands de l'Association d'autorégulation volontaire des services multimédias (*Verhaltenssubkodex für Suchmaschinenanbieter - « VK-S » - der Freiwillige Selbstkontrolle Multimedia-Diensteanbieter*) du 21 avril 2004, disponible en anglais sur http://www.fsm.de/en/Subcode_of_Conduct_for_Search_Engine_Providers (consulté le 1^{er} juillet 2008) et en allemand (version originale) : http://www.fsm.de/de/Subkodex_Suchmaschinenanbieter

volontariat, des mesures telles que la mise en place d'initiatives d'autorégulation, l'adoption d'un code de conduite et la gestion d'un bureau des réclamations (comme c'est le cas avec FSM) servent à externaliser la responsabilité. De cette façon, l'exigence de conformité semble provenir d'une instance extérieure, que les parties elles-mêmes ressentent comme une pression externe. Cependant, ceci ne signifie pas qu'il existe pour autant une obligation légale effective.

3.3. Les besoins réels en matière de réglementation

En dernière analyse, cependant, les participants sont parvenus à la conclusion que l'existence même d'un besoin de réglementation de la part des moteurs de recherche n'était pas un fait acquis. Au vu de la liste détaillée de Wolfgang Schulz sur les risques liés aux moteurs de recherche³², l'un des participants a objecté que cette approche était excessivement alarmiste. Il s'est interrogé, par exemple, sur la signification précise de la fragmentation de la sphère publique et des dangers que cela représente. La réponse a fait état des études qui ont révélé que l'utilisation des « médias de recherche » limitait l'éventail des thèmes sur lesquels un individu collecte des informations. En contrepartie, les médias traditionnels placent les lecteurs face à des thèmes qu'ils n'auraient pas forcément sélectionnés eux-mêmes mais qui sont néanmoins susceptibles de les intéresser ou de répondre à leurs besoins³³. En réponse, l'argument a été avancé que les médias de recherche pourraient être considérés comme un facteur facilitant l'émergence de communautés d'intérêts communs, ce qui, par conséquent, renforce la cohésion sociale.

Les participants étaient unanimes pour convenir que la façon dont les gens appréhendent l'information est en train de changer radicalement. Jane Buckingham, fondatrice de l'Intelligence Group, a rapporté le cas d'un étudiant qui affirmait : « Si une information est importante, elle saura où me trouver³⁴. » Certains participants pensent effectivement que le problème des moteurs de recherche peut être lié au fait que, pour le moment, nous nous trouvons dans une phase de transition : l'impact et les répercussions des nouveaux médias et les pièges qu'ils peuvent receler sont encore flous. Dans la même optique, l'idée a été lancée que l'utilisation des moteurs de recherche pourrait, en fait, contribuer à limiter globalement les besoins d'une réglementation. Par exemple, dans le domaine de la protection des consommateurs, les moteurs de recherche peuvent aider à renforcer la position des utilisateurs en facilitant l'accès à une meilleure connaissance des produits qui leur sont proposés. De même, en matière de pluralisme des médias, les moteurs de recherche ont la capacité de guider leurs utilisateurs vers une offre pluraliste de services. Par conséquent, la transparence s'en trouve renforcée et les besoins en matière de réglementation déclinent. Néanmoins, pour réels qu'ils soient, ces avantages ne doivent pas masquer la nécessité de traiter la question spécifique des raisons éventuelles en faveur d'une intervention réglementaire présentées par les moteurs de recherche eux-mêmes.

Par conséquent, d'autres participants considèrent la liste des risques susmentionnée comme une vision optimiste et non alarmiste. Selon eux, la plupart des risques répertoriés sont déjà couverts par la législation actuelle, par l'autorégulation ou par l'auto-organisation des utilisateurs. Ainsi, le risque de voir une distorsion de la concurrence est une question qui relève du droit de la concurrence et l'accès aux contenus préjudiciables peut être restreint par une pratique d'autorégulation sur la base de codes de conduite. Enfin, les phénomènes tels que la fragmentation de la sphère publique peuvent être compensés par la capacité d'Internet à fédérer des groupes d'intérêts communs. En conclusion, les participants ont suggéré que les législateurs pourraient utiliser leur temps et leurs efforts de façon beaucoup plus fructueuse si, en lien avec les radiodiffuseurs et toutes les ressources associées (telles que les EPG) qui sont disponibles sur une plateforme donnée, ils leurs demandaient de fournir des informations complètes sur leurs programmes correctement marquées. De cette façon, les informations auxquelles ont accès les moteurs de recherche audiovisuels seraient correctes et il n'y aurait pas lieu de réclamer une réglementation *ex ante* de ces mêmes outils.

32) Cette liste englobe les risques suivants : accès à des contenus préjudiciables, accès à des contenus illicites, discrimination des contenus, tromperie à l'égard des utilisateurs, exercice d'une influence sur l'opinion, fragmentation de la sphère publique, exploitation d'œuvres protégées, exploitation de données personnelles, distorsion de la concurrence, y compris transfert de la force du marché sur d'autres marchés (par ex. la publicité). Voir l'article de Wolfgang Schulz dans cette publication.

33) Klaus Schönbach et autres, *Online Newspapers: A Substitute for Print Newspapers and Other Information Channels?* 6th World Media Economics Conference, Centre d'études sur les médias et Journal of Media Economics, HEC Montréal, Montréal, Canada, 12-15 mai 2004, disponible sur : <http://cf.uba.uva.nl/nl/handl/googlescholar/>, consulté le 9 juillet 2008.

34) Brian Stelter, *Finding Political News Online, the Young Pass It On*, disponible sur : http://www.nytimes.com/2008/03/27/us/politics/27voters.html?_r=1&adxnln=1&oref=slogin&adxnlnx=1215003699-5hhJsU3ewQ+YzGMGjWsk0Q, consulté le 2 juillet 2008.

Les défis technico-économiques liés aux moteurs de recherche audiovisuels

Ramón Compañó ()*

*Commission européenne – Centre commun de recherche (CCR)
Institut des études de prospective technologique (IPTS)*

Remerciements :

Ce document est dédié à mon collègue et ami Boris Rotenberg, décédé le 23 décembre 2007 dans un tragique accident de ski. Bon nombre d'idées exposées ici ont été développées dans le cadre de nos discussions sur les perspectives des moteurs de recherche. Boris Rotenberg restera toujours présent dans ma mémoire comme un chercheur dynamique et brillant, et un homme plein d'enthousiasme.

Cette étude a reçu le soutien de la Commission européenne dans le cadre du projet « Chorus » (numéro de contrat 2006-045480).

(*) Avertissement

Les opinions exprimées dans cette publication relèvent de la responsabilité exclusive des auteurs et ne reflètent pas nécessairement le point de vue de la Commission européenne. Ni la Commission européenne, ni les représentants de la Commission ne peuvent être tenus responsables de l'usage qui est fait des informations présentées.

Introduction

Les moteurs de recherche audiovisuels (AV) sont en passe de devenir des outils indispensables dans l'univers audiovisuel, comparables à ce que furent les recherches dans le texte pour l'environnement numérique textuel actuel. Les applications de recherche AV nous permettront de répertorier, de filtrer et de référencer en toute fiabilité toutes formes de création audiovisuelle (individuelle ou collective). En outre, la recherche AV jouera un rôle central dans les applications de partage de fichiers essentiellement audiovisuels. Elle va également entraîner l'émergence de nouveaux modes de traitement de l'information numérique. La technologie de reconnaissance des formes, par exemple, permettra aux utilisateurs de rechercher des catégories d'images ou de séquences filmées. De même, la recherche AV pourrait permettre de collecter toutes les conversations voix sur IP passées, au cours desquelles certains mots clefs ont été utilisés.

Néanmoins, pour que ce type d'applications puisse émerger, la technologie de recherche doit se développer rapidement, tant du point de vue de sa portée que de sa nature. Les efforts devront porter sur le développement de nouvelles techniques de recherche AV, par exemple à partir de notre comportement. La recherche AV est donc classée parmi les priorités absolues des trois principaux opérateurs américains de moteurs de recherche Google, Yahoo! et Microsoft. En Europe, la mise en place de l'initiative française Quaero¹ pour le développement d'un portail de recherche AV ultra-performant ou du programme de recherche allemand Theseus² est une preuve supplémentaire de l'importance accordée à la recherche AV.

Il existe toute une variété de moteurs de recherche AV pour différents types d'applications. Ils se distinguent par le type de contenu audiovisuel (texte, audio, vidéo), la plateforme technologique (téléphone portable, PC) ou l'application (par exemple les informations ouvertes au public sur le Web contre les bases de données propriétaires). Du point de vue du grand public, les plus connus sont les « moteurs de recherche sur le Web ». Le terme « moteurs de recherche sur le Web » désigne un service disponible sur l'Internet public permettant aux internautes de chercher et de retrouver des contenus ou des informations³ à partir de l'Internet accessible au public⁴. Google, Yahoo!, Microsoft et le service de moteur de recherche d'AOL figurent parmi les plus connus. Les moteurs de recherche sur le Web doivent être différenciés des moteurs qui recherchent des informations à partir de sources non accessibles au public. Parmi ces derniers, citons, à titre d'exemple, les moteurs dédiés exclusivement à la recherche d'informations dans de vastes bases de données propriétaires internes aux entreprises (par exemple ceux qui recherchent des produits sur eBay ou Amazon, ou qui procèdent à la recherche d'informations sur Wikipedia), ou les moteurs qui recherchent des informations non accessibles aux moteurs de recherche sur le Web⁵. De la même façon, nous excluons de cette définition les moteurs qui recherchent des données à partir de réseaux fermés *peer-to-peer* ou d'applications non accessibles au public et qui n'effectuent pas leurs recherches sur l'Internet accessible au public. Ils peuvent servir à des recherches d'ordre général, tels que Google, Yahoo! ou MS Live Search, ou être spécialisés sur des thèmes particuliers. Les portails de recherche thématique, ou verticale, peuvent être dédiés spécialement aux enfants⁶, à la recherche de personnes⁷, spécialisés sur les questions de santé⁸, la

1) <http://www.quaero.org/> (consulté le 6 juin 2008).

2) <http://theseus-programm.de/> (consulté le 6 juin 2008).

3) Dans un souci de précision, j'emploie le terme « contenu » ou « information » pour désigner le résultat d'une recherche plutôt que « page Web », car de nombreux moteurs recherchent des informations autres que des pages Web, notamment des fichiers musique, livres numériques, codes logiciels ou autres.

4) Pour une définition similaire, voir James Grimmelmann, *The Structure of Search Engine Law* (draft), 13 octobre 2006, p. 3, sur http://works.bepress.com/james_grimmelmann/13/ (consulté le 6 juin 2008). Il est établi que de nombreux constats du présent rapport peuvent s'appliquer à différents types de moteurs de recherche.

5) Une partie du Web accessible au public ne peut pas être détectée par les moteurs de recherche Web car les programmes automatisés des moteurs de recherche qui indexent le Web, les robots, *crawlers* ou *spiders* n'y ont pas accès du fait du caractère dynamique du lien ou parce que l'information est protégée par des dispositifs de sécurité. Bien que la technologie des moteurs de recherche se perfectionne avec le temps, parallèlement, le nombre de pages Web augmente massivement, ce qui empêchera vraisemblablement le Web « invisible » ou « profond » de disparaître dans un proche avenir. En date du mois de mars 2007, le nombre de pages Web (par opposition aux sites) est estimé entre 15 et 30 millions, dont un quart à un cinquième est accessible aux moteurs de recherche. Voir à ce sujet <http://www.pandia.com/sew/383-web-size.html> (consulté le 6 juin 2008) et <http://technology.guardian.co.uk/online/story/0,,547140,00.html> (consulté le 6 juin 2008).

6) Par ex. Fragfinn (<http://www.fragfinn.de/>).

7) Par ex. Spock (<http://www.spock.com/>).

8) Par ex. Healia (<http://www.healia.com/>).

recherche de codes logiciels⁹, la recherche de stations de radios, de musique, spécialisés dans les sciences, etc. Une liste régulièrement mise à jour des moteurs de recherche classés par thème est disponible sur Pandia¹⁰.

Contrairement aux moteurs de recherche qui explorent les informations accessibles au public, certains moteurs font des recherches sur des *ordinateurs* personnels ou dans des *bases de données multimédia propriétaires fermées*. Les premiers sont largement répandus et peuvent se télécharger à partir de la gamme d'outils proposée gratuitement par les fournisseurs des principaux moteurs de recherche. Ils permettent de rechercher des informations sur un ordinateur personnel. En dépit de certaines réserves concernant la confidentialité des données, ces outils sont très appréciés, car ils sont gratuits. Les entreprises qui gèrent des données sensibles ont recours à des solutions métiers qui tiennent compte de leurs besoins spécifiques, comme, par exemple, les bibliothèques nationales. Une autre forme d'outil concerne *la recherche sur réseaux peer-to-peer*, qui vise à retrouver des contenus audiovisuels sur les sites de PC privés (comme, par exemple, les outils de recherche opérant sur les plateformes d'échange BitTorrent). Enfin, une nouvelle tendance émerge avec l'apparition de la *recherche mobile*.

Le chapitre suivant est consacré à la présentation de certains aspects technologiques élémentaires afin de mieux comprendre les principaux défis technologiques.

1. Les défis technologiques

1.1. Les éléments de base du moteur de recherche

Le niveau d'avancement technologique des moteurs de recherche dépend de l'outil de recherche spécifique (audio, image ou vidéo) et de la plateforme (par ex. architecture répartie entre plusieurs ordinateurs, réseau *peer-to-peer*, etc.). Plutôt que de se livrer à une description exhaustive, ce chapitre se propose de présenter les différents éléments technologiques pour mieux comprendre les principaux défis technologiques.



Figure 1 : Processus d'un moteur de recherche

En résumé, un moteur de recherche est composé d'un certain nombre de composants techniques qui exécutent quatre opérations essentielles avec les informations : collecte et indexation des informations, requête de l'utilisateur, restitution de l'information demandée et accès de l'utilisateur à l'information. La collecte des informations est assurée par des agents logiciels automatisés appelés robots, *spiders* ou *crawlers*¹¹.

Une fois que le robot a téléchargé et stocké une page sur le propre serveur du moteur de recherche, un second programme, dit d'indexation, en extrait divers éléments d'information. Les principaux facteurs pris en compte portent sur les mots contenus par la page Web, l'emplacement de ces mots-clefs (par ex. le titre), la valeur attribuée à certains mots spécifiques (les mots rares, par exemple, ont une valeur de recherche supérieure aux mots d'utilisation courante), la proximité des mots ou les hyperliens contenus sur la page. Il est important de noter que l'index n'est pas une reproduction de la page en tant que telle, ni un document qu'un internaute pourrait lire. L'index est soumis à une analyse complémentaire et un référencement croisé pour former l'index du module d'exécution utilisé pour l'interaction avec l'internaute.

⁹ Par ex. Koders (<http://www.koders.com/>) ou Krugle (<http://www.krugle.org/>).

¹⁰ <http://www.pandia.com/powersearch/index.html> (consulté le 6 juin 2008).

¹¹ Il existe également des alternatives non automatisées ou semi-automatisées sur le marché, telles que le projet de répertoire ouvert où le Web est catalogué par les utilisateurs, ou les moteurs de recherche qui puisent dans les connaissances du public pour fournir les informations pertinentes à leurs utilisateurs, tels que Wiki Search, l'initiative de Wikipedia en matière de moteur de recherche. (http://search.wikia.com/wiki/Search_Wikia) (consulté le 6 juin 2008), ou ChaCha (<http://www.chacha.com/>) (consulté le 6 juin 2008). Voir Wade Roush, *New Search Tool Uses Human Guides*, *Technology Review*, 2 février 2007, sur <http://www.techreview.com/Infotech/18132> (consulté le 6 juin 2008).

Lorsqu'un internaute envoie une requête, le moteur explore l'index du serveur. L'algorithme de la requête est « l'âme » du moteur de recherche. L'algorithme prend en charge deux processus clefs. Dans un premier temps, il définit le processus de correspondance entre la requête de l'internaute et le contenu de l'index. Ensuite, il trie et classe les différentes occurrences. Par conséquent, le degré de pertinence pour l'internaute dépend de l'algorithme.

L'internaute reçoit les contenus sous forme de lien vers la source originale ou à partir du cache du moteur de recherche, c'est-à-dire du « fichier provisoire » provisoirement archivé sur le propre serveur du moteur de recherche. Le « cache » est une copie de la dernière visite effectuée par le robot du moteur de recherche sur cette page. Par conséquent il n'est pas forcément à jour mais peut s'avérer utile pour l'internaute si le serveur ou la page sont provisoirement indisponibles ou si l'internaute souhaite connaître les dernières modifications effectuées sur cette page.

Les moteurs de recherche courants sont le plus souvent textuels, même pour les contenus audiovisuels. Cela signifie que les contenus non-textuels tels que les fichiers images, audio ou vidéo sont indexés, comparés et classés en fonction de propriétés textuelles telles que le nom des fichiers, les tags (étiquettes), le texte accompagnant des images ou des vidéos (par exemple les légendes), et même le texte d'ancrage des liens qui mènent directement vers le contenu AV. À titre d'exemple, nous citerons Truevo pour les clips video¹² et SingingFish pour les fichiers audio¹³.

Alors que la recherche textuelle est efficace pour les fichiers contenant exclusivement du texte, cette technique présente de gros inconvénients lorsqu'il s'agit de retrouver des informations à partir de contenus autres que du texte. Ainsi, des images très pertinentes pour l'objet de la requête ne seront pas répertoriées par le moteur de recherche si le fichier n'est pas accompagné des tags correspondants ou des indices textuels. Par exemple, même si une vidéo comporte une montagne rouge, le moteur de recherche ne retiendra pas cette vidéo dans le cadre d'une requête à partir du mot-clef « montagne rouge ». Il en va de même pour toute autre information produite sous une forme non-textuelle. Autrement dit, un nombre considérable d'informations pertinentes est systématiquement exclu du classement des moteurs de recherche et reste inaccessible à l'utilisateur. En retour, ce phénomène affecte également la production de toutes sortes de nouvelles informations¹⁴.

Il y a donc une énorme lacune dans notre processus de recherche des informations. Cette lacune se creuse en même temps que se développe le volume des informations non-textuelles produites. Les chercheurs du monde entier s'efforcent de combler cette lacune. L'un des axes de développement porte sur l'amélioration de la production de métadonnées qui décrivent le contenu audiovisuel au format texte. Ainsi, une solution consisterait à développer un logiciel « intelligent » qui étiquette automatiquement le contenu audiovisuel¹⁵. Néanmoins, même si la technologie évolue, l'étiquetage automatique reste très inefficace en raison de la complexité des algorithmes, des exigences requises au niveau du traitement informatique et des procédures de calcul. Une autre possibilité consisterait à créer un système qui étiquette les images en associant une vision artificielle aux éléments saisis par l'utilisateur¹⁶. Cependant, l'étiquetage manuel est une tâche fastidieuse et, par conséquent, coûteuse.

1.2. Les défis

Les défis technologiques sont multiples et complexes, car ils concernent à la fois les principes théoriques fondamentaux et les moyens de mise en œuvre. Dans un souci de simplification, nous regrouperons ces défis sur le modèle de la figure 1, à savoir collecter un maximum d'information (collecte des informations), comprendre les requêtes des utilisateurs, s'adapter au contexte de l'utilisateur (interaction avec l'utilisateur), fournir les résultats les plus pertinents, fournir des résultats qui soutiennent l'interaction de l'utilisateur (restitution des résultats). Un autre défi

12) <http://www.truevo.com/> (consulté le 6 juin 2008).

13) SingingFish a été racheté AOL en 2003 et a cessé d'exister en tant que service distinct en 2007. Voir <http://en.wikipedia.org/wiki/Singingfish> (consulté le 6 juin 2008).

14) Voir Matt Rand, *Google Video's Achilles' Heel*, Forbes.com, 10 mars 2006, sur http://www.forbes.com/2006/03/10/google-video-search-tveyes-in_mr_bow0313_inl.html (consulté le 6 juin 2008).

15) À ce sujet, voir James Lee, *Software Learns to Tag Photos*, Technology Review, 9 novembre 2006, sur <http://www.technologyreview.com/Infotech/17772/>. Voir Chris Sherman, *Teaching Google to See Images*, Search Engine Land, 5 avril 2007, sur <http://searchengineland.com/070405-172235.php> (consulté le 6 juin 2008).

16) Voir Michael Arrington, *Polar Rose: Europe's Entrant Into Facial Recognition*, Techcrunch, 19 décembre 2006, sur <http://www.techcrunch.com/2006/12/19/polar-rose-europes-entrant-into-facial-recognition> (consulté le 6 juin 2008).

concerne la méthode de gestion de la multiplicité et de l'interopérabilité des dispositifs et des plateformes. Ce point sera brièvement abordé dans le chapitre suivant.

1.2.1. Collecte des informations : gérer l'explosion des données

Le volume des informations créées, stockées et répliquées était de 3 milliards de giga-octets (3 exa-octets) en 2000, 24 exa-octets en 2003 et 161 exa-octets en 2008¹⁷. À titre de comparaison, ce dernier chiffre représente l'équivalent de trois millions de fois le volume d'informations contenues dans tous les livres jamais écrits jusqu'à présent. Ce chiffre devrait atteindre 988 milliards de giga-octets d'ici 2010¹⁸. Nous sommes incontestablement en présence d'une explosion du volume de données numériques.

Cette masse gigantesque de données se présente sous des formats très divers et les contenus ont évolué largement au-delà de la simple description textuelle. En fait, même si les données étaient encore essentiellement textuelles au début de l'ère numérique, on pense qu'actuellement, 93 % des données sur Internet sont sous format multimédia et cette proportion avoisinera les 99 % en 2010. Par ailleurs, il est intéressant de noter que 70 % de ces données sont créés ou réalisés par des utilisateurs non-professionnels¹⁹. Ceci met en lumière le fait qu'au cours des prochaines années, les utilisateurs vont progressivement passer du statut de simples consommateurs à celui de producteurs, fournisseurs et médiateurs de contenus. Il est légitime de supposer que pour être en mesure de traiter le volume croissant de contenus audiovisuels, les moteurs de recherche vont devoir se transformer en moteurs de recherche AV. Les futurs moteurs de recherche audiovisuels devront affronter deux problèmes majeurs. D'une part, ils devront être évolutifs pour pouvoir traiter un nombre sans cesse croissant de données. Le volume de données résulte non seulement de la création d'une profusion de données originales à stocker, comme nous l'avons mentionné ci-dessus, mais aussi, et surtout, de l'énorme volume de données de traitement et de métadonnées annexes nécessaires pour assurer une recherche efficace de contenus AV. En fait, l'indexation des fichiers audio, images ou vidéo exige des ressources considérables en termes de puissance de traitement informatique et de capacité de stockage. D'autre part, ils devront être en mesure de traiter différents types de matériel audiovisuel. Cette exigence impliquera sans doute l'utilisation de concepts de recherche moins déterministes, c'est-à-dire le passage d'un niveau de correspondance exact à un niveau de correspondance approximatif. L'exigence d'une évolutivité croissante et d'une moindre détermination des contenus va entraîner l'apparition de nouveaux types d'architecture informatique et de concepts de recherche. Ces derniers sont représentés dans le schéma ci-dessous et feront l'objet d'un point au chapitre suivant.

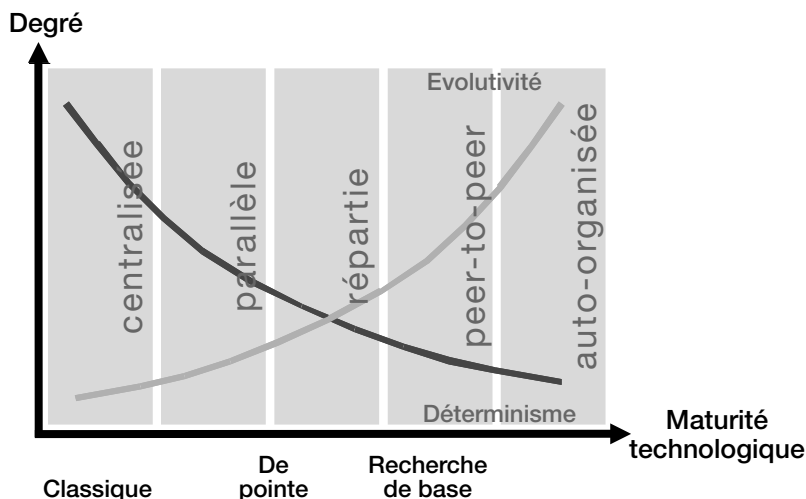


Figure 2

17) Voir à titre comparatif : <http://www.pandia.com/sew/383-web-size.html> (consulté le 6 juin 2008) et <http://technology.guardian.co.uk/online/story/0,,547140,00.html> (consulté le 6 juin 2008).

18) Voir Andy McCue, *Businesses face data 'explosion'*, ZDNet, 23 mai 2007, sur : <http://news.zdnet.com.co.uk/itmanagement/0,1000000308,39287196,00.htm> (consulté le 6 juin 2008), en référence à l'étude d'IDC/EMC *The expanding Digital Universe*.

19) Fauto Rabitti, présentation de Sapir au colloque de Vilamoura (Faro, Portugal), 16 avril 2008.

Les contenus multimédia créés quotidiennement par les internautes vont dépasser la capacité de traitement des moteurs de recherche actuels pour l'indexation et la recherche de contenus AV au sein de textes associés et de métadonnées. Ce phénomène aura des répercussions sur l'architecture informatique (depuis la configuration centralisée jusqu'aux futurs groupes auto-organisés) et les modes de recherche (plus ou moins déterministes, par exemple avec des concordances exactes ou approximatives, ou des recherches précises ou approchantes)²⁰.

Au début, les contenus étaient stockés sur des serveurs centralisés. Tant que le volume des informations était raisonnable et les contenus essentiellement textuels, la recherche était raisonnablement efficace. Lorsque la masse de contenus a augmenté, les systèmes informatiques centralisés ont cessé d'être en mesure de répondre aux utilisateurs dans des délais raisonnables. Il a fallu remplacer les systèmes centralisés dans un premier temps par des architectures parallèles et, par la suite, par des structures réparties.

En fait, de nombreux scientifiques pensent que les systèmes de traitement les plus évolués en architecture répartie ne seront pas suffisants, à l'avenir, pour faire face au volume croissant de données. Ils pensent que les futurs moteurs de recherche sur le Web fonctionneront selon un système *peer-to-peer* et, qu'à terme, compte tenu de la complexité du fonctionnement de ces architectures informatiques *peer-to-peer* pour obtenir des résultats efficaces, il sera nécessaire d'adopter un certain niveau d'auto-organisation pour gérer cette complexité de façon autonome.

Actuellement, les systèmes répartis constituent l'architecture informatique la plus évoluée pour les moteurs de recherche. Les systèmes *peer-to-peer* sont parvenus à un niveau de développement avancé, plusieurs prototypes de recherche ont démontré leur potentiel²¹ et certaines entreprises, telles Faroo²², proposent d'ores et déjà des versions Bêta de moteurs de recherche *peer-to-peer* sur le Web avec un système d'index, d'exploration et de classement réparti entre les utilisateurs. La recherche *peer-to-peer* pourrait offrir une alternative évolutive aux moteurs de recherche en grappe si les techniques de pointe appliquées dans ce domaine prouvent leur fiabilité avec un grand nombre de postes en réseau, c'est-à-dire si elles sont mises en œuvre par une vaste communauté d'utilisateurs. Pour la recherche plein texte, les configurations *peer-to-peer* devraient contribuer à la segmentation des documents (réseau dédié de recherche non structurée, par exemple Gnutella) et à la segmentation textuelle (réseau dédié de recherche structurée, par exemple Chord, P-Grid), alors que pour la recherche audiovisuelle, le *peer-to-peer* va essentiellement contribuer aux tâches d'indexation ainsi qu'à la recherche de similarités. Il reste à savoir quand, en admettant que ce soit le cas un jour, les architectures *peer-to-peer* seront prêtes à remplacer les systèmes répartis. Certaines solutions (isolées) de recherche *peer-to-peer* de similarités au niveau des propriétés AV individuelles (par ex. la couleur ou la forme) sont d'ores et déjà disponibles, mais les systèmes combinant texte et propriétés multiples (par ex. couleur ET forme) ne sont pas encore opérationnels.

D'un point de vue opérationnel, les futurs moteurs de recherche devraient être plus performants lorsqu'ils accèdent au Web caché et recherchent toutes sortes de contenus audiovisuels. Le volume croissant des données pose un véritable défi en ce qui concerne le stockage et le traitement des données, en particulier pour l'indexation et le délai de réponse aux requêtes des utilisateurs. Il est important de noter que les formats multimédia (images, vidéo, audio, 3D, etc.) rendent l'indexation beaucoup plus complexe que les formats texte. À moins que de nouvelles architectures se mettent en place et que des solutions soient disponibles pour les contenus non-indexés, on risque de voir se creuser une large brèche dans la recherche d'information AV. Il est probable que les moteurs de recherche s'appuieront pendant une longue période sur deux approches complémentaires de recherche AV, à savoir les technologies basées sur l'annotation, où les indices sont générés automatiquement ou manuellement (par exemple l'étiquetage), et les technologies basées sur le contenu, par exemple avec la recherche de similarités.

Pour optimiser le processus de recherche, à un stade ultérieur, il serait judicieux de procéder à l'indexation au niveau même de la source. Ainsi, les systèmes commerciaux auraient une série de données déjà intégrées, telles que l'auteur, le processus de production, les données géographiques, etc. Parallèlement, l'utilisateur aurait la possibilité d'indexer, d'ajouter une étiquette, un commentaire, etc. Ces métadonnées associées aux contenus AV pourraient faciliter la recherche de contenus pertinents.

20) <http://www.sapir.eu/papers/sapir-poster.pdf> (consulté le 6 juin 2008).

21) <http://www.sapir.eu/> (consulté le 6 juin 2008).

22) <http://www.faroo.com/> (consulté le 6 juin 2008).

1.2.2. Comprendre les requêtes des utilisateurs : combler le fossé sémantique

La saisie de mots-clés n'est pas le mode d'interaction idéal pour la recherche d'informations. En se rapprochant d'une sorte de langage plus naturel, on pourrait remédier à ce « jargon de mots-clés », mais cette solution n'est pas près de voir le jour. Alors qu'à l'heure actuelle, il est possible d'effectuer une recherche à partir d'une « phrase simple », dès que le niveau de complexité augmente un tant soit peu, les moteurs ont du mal à traiter la requête, car la recherche contextuelle n'est pas encore possible. Si nous saisissons le mot « jaguar », par exemple, le moteur de recherche sera incapable de distinguer s'il s'agit de la marque automobile ou de l'animal, puisque le traitement de la requête est lié à la capacité du système à comprendre la sémantique d'une requête. Comme nous l'avons mentionné plus haut, les progrès réalisés dans le domaine de l'apprentissage automatique et de l'intelligence artificielle sont considérables, néanmoins il reste un long chemin à parcourir avant de voir se concrétiser le Web sémantique.

Les performances de la recherche baissent proportionnellement à l'augmentation du niveau de « complexité » des contenus, comme l'illustre l'exemple des images. Du point de vue de la recherche, les images les plus simples sont celles qui présentent une grande clarté au niveau des contours, des couleurs, de la géométrie (par ex. sans déformation), des formes, etc. C'est le cas, par exemple, des images « artificielles », telles que les logos ou les symboles. Ces images sont relativement faciles à trouver avec des moteurs de recherche, car les méthodes basées sur la similarité fonctionnent très bien. Par conséquent, les moteurs de recherche sont utilisés dans des applications de repérage des contrefaçons de marques, puisque les marques commerciales sont souvent représentées par des images aux caractéristiques brutes.

Les images présentant des mesures de similarité spécifiques constituent la catégorie suivante. Elles sont plus complexes que les images « artificielles » mentionnées précédemment, mais elles comportent des caractéristiques spécifiques qui permettent de les retrouver rapidement. Prenons, par exemple, des images représentant des visages vus de face²³ : leurs caractéristiques individuelles (yeux, bouche et nez) sont clairement définies, de même que la relation spatiale entre ces caractéristiques. Un autre exemple est fourni par les empreintes digitales. Le contraste des couleurs est marqué et la similarité des formes concordantes est établie en les comparant avec des empreintes déjà stockées sous forme d'images dans une base de données.



Figure 3 : images « semi-artificielles »

Les images semi-artificielles sont plus complexes du point de vue de la recherche. Celles-ci sont des images présentées dans des conditions bien définies, avec, notamment, un seul objet par image apparaissant sur un fond homogène. La série de poissons et de bracelets de la figure 3 est conforme à ces critères. La recherche de ce type d'images est utile dans le cadre de projets pédagogiques, en particulier pour les encyclopédies, ainsi que de certaines applications commerciales (par ex. pour rechercher des produits spécifiques)²⁴. Les images les plus difficiles à rechercher sont les images réelles,

23) Le logiciel est capable de reconnaître des visages sur la base des zones fortement contrastées telles que les yeux, le nez et la bouche. Dans des conditions optimales (vue de face en gros plan, cliché de qualité supérieure) les systèmes atteignent 95 % de précision avec un minimum de faux positifs. Pour les vues prises de côté, des images de qualité médiocre ou des images anciennes issues de collections du patrimoine culturel, les résultats restent bien supérieurs à la moyenne. Pour un aperçu général, voir Yang, M.H., Kriegman, D.J., and Ahuja. N. 2002. *Detecting Faces in Images: A Survey. IEEE Transactions on Pattern Analysis and Machine Intelligence* 24(1), 34-58.

24) Par ex. <http://www.riya.com/> (consulté le 6 juin 2008).

telles que celles qui composent les albums photos ou que l'on trouve sur Internet. Ces images comportent de multiples catégories d'éléments (par exemple paysages, personnages, art, objets artificiels, etc.), ce qui les rend « versatiles » et complexes, et, par conséquent, difficiles à analyser et à traiter.

Technologiquement parlant, les moteurs peuvent retrouver assez facilement des images aux caractéristiques brutes, telles que les images artificielles, en analysant les éléments de base tels que la couleur, la texture et la forme de l'objet. Par conséquent, les moteurs de recherche travaillant à partir de la similarité se révèlent efficaces dans certains cas particuliers, notamment la recherche dans des bases de données de marques commerciales²⁵, la détection de contrefaçons^{26,27}, la recherche de séquences vidéos ayant des contenus visuels similaires et la recherche de musiques au rythme comparable. Pour toutes ces applications, les caractéristiques brutes telles que la couleur et la texture pour les images et la vidéo, ou le rythme, la mélodie ou la gamme de fréquences dominantes pour la musique, sont étroitement liées aux objectifs de recherche spécifiques de chaque application.

Ainsi, dans SongTapper²⁸, un utilisateur peut rechercher une chanson en tapant le rythme de cette chanson sur la barre d'espace du clavier, ou sur Midomi²⁹, en chantonnant la chanson au micro.

Les caractéristiques plus conceptuelles sont beaucoup plus complexes, puisqu'elles font souvent intervenir la nécessité de « comprendre » le contenu. Ainsi, les humains sont capables d'identifier sur le champ une projection bidimensionnelle d'un objet tridimensionnel (3D), c'est-à-dire que nous pouvons facilement reconnaître une image différente comme étant simplement une autre projection (sous un angle différent) d'un seul et même objet. En revanche, les moteurs de recherche sont incapables de reconnaître la représentation bidimensionnelle d'un objet 3D. Ils n'ont pas la capacité de comprendre le contenu d'une image, ni du mode de description d'un objet par un être humain. La différence entre deux descriptions d'un objet par différentes représentations constitue le fossé sémantique. *Comblé ce fossé sémantique constitue l'un des défis majeurs de l'informatique.*

Il ne s'agit pas seulement d'un défi académique, car le fait de combler ce fossé sémantique permettrait, en pratique, de développer des interfaces plus conviviales. Celles-ci seraient en mesure de traduire un savoir contextuel exprimé en langage naturel (conceptuel) en une opération élémentaire et reproductible (brute) sur un ordinateur. Le développement de ce type de moteurs de recherche convivial permettrait de rendre accessible aux citoyens la multitude des ressources multimédia stockées dans les bibliothèques, les bases de données et les collections. Ces systèmes seraient en mesure de comprendre le contenu sémantique d'une requête et pas simplement les caractéristiques informatiques brutes qui la sous-tendent. ImageScape³⁰ fut l'un des premiers systèmes de recherche basé sur le contenu à intégrer des dispositifs visant à combler le fossé sémantique au niveau de l'interface de requête, de l'indexation et de la restitution des résultats. Ce moteur de recherche précoce permettait à l'utilisateur de rechercher de multiples objets visuels tels que skis, arbres, eau, etc., à l'aide d'icônes intégrées dans un index contenant plus de 10 millions d'images et de vidéos. La détection automatique de tous les contenus sémantiques au sein d'une image telle qu'un visage, un arbre, un animal, etc., en présence d'arrière-plans complexes, constitue à l'heure actuelle un énorme défi. Les chercheurs qui travaillent sur le classement d'images complètes buttent souvent sur les obstacles technologiques et sur les questions pratiques telles qu'une granularité picturale trop forte pour être utile dans des applications concrètes.

En dépit du fait que le problème du fossé sémantique est loin d'être résolu, la recherche progresse en permanence. Actuellement, les méthodes axées sur l'apprentissage de la sémantique sont nombreuses et englobent, entre autres, des concepts tels que l'examen des associations cachées durant l'indexation des images, le développement de dictionnaires visuels regroupant des couleurs et des textures similaires, ou l'introduction de méthodes d'apprentissage.

25) Eakins, J.P., Riley, K.J., and Edwards, J.D. 2003. *Shape Feature Matching for Trademark Image Retrieval*. CIVR, 28-38.

26) Jaimes, A and Chang, S-F. 2002 *Duplicate Detection in Consumer Photography and News Video*, Conférence internationale ACMsur le multimédia, 423-424.

27) Joly, A., Buisson, O., and Frelicot, C. *Robust content-based copy detection in large reference database*, Conférence internationale sur la recherche d'images et de vidéos, 2003.

28) <http://www.songtapper.com/> (consulté le 6 juin 2008).

29) <http://www.midomi.com/> (consulté le 6 juin 2008).

30) Lew, M.S. 2000. *Next Generation Web Searches for Visual Content*. *IEEE Computer*, novembre, 46-53.

1.2.3. Restituer des résultats plus pertinents : intégrer l'utilisateur dans la boucle

Le degré de pertinence dépend de la qualité de l'algorithme de recherche. L'une des difficultés rencontrées par les développeurs de moteurs de recherche concerne la pollution et les facteurs externes négatifs (le *référencement abusif* ou le *link-bombing* sont des exemples bien connus d'outils non autorisés d'optimisation des moteurs de recherche auxquels les développeurs de moteurs de recherche se sont trouvés confrontés).

Une solution permettant d'augmenter la qualité des systèmes de recherche par similarité consiste à leur permettre d'apprendre à l'usage, en intégrant un retour systématique de la requête de l'utilisateur. Ce processus interactif est appelé « pertinence », « affinement de la recherche » ou « recherche interactive ». Une liste d'images candidates est présentée à l'utilisateur qui est invité à classer la pertinence de ces images par rapport à sa requête. L'un des problèmes majeurs du retour de pertinence consiste à pouvoir intégrer de brèves sessions de formation, car l'utilisateur type aura tendance à marquer un nombre très insuffisant d'images par rapport aux besoins de l'algorithme. Pour optimiser le retour de l'utilisateur, une option consiste à proposer une représentation appropriée de l'information recherchée. Certains moteurs de recherche commencent à introduire des interfaces visuelles dans la structure de l'information et à rendre visible le lien vers d'autres éléments de l'information. Clusty³¹ offre un exemple de moteur de recherche qui regroupe les résultats, et KartOO³² propose une interface visuelle des liens et des relations entre les résultats trouvés. Une autre méthode pour améliorer la pertinence consiste à utiliser un système de *bookmarking* social tel que Mister-Wong³³, ou d'intégrer le retour des utilisateurs dans le processus de recherche, à l'instar du moteur de recherche ChaCha³⁴.

Les services de recherche seraient plus efficaces et plus pertinents si les requêtes éventuelles pouvaient être anticipées, du moins dans une certaine mesure. Cela implique la personnalisation et l'adaptation du moteur de recherche en fonction de l'utilisateur. Dans les cas extrêmes, un niveau avancé de personnalisation implique la nécessité, pour les systèmes, de stocker un volume équivalent d'informations personnelles sur les individus. Par conséquent, la somme de toutes les requêtes des utilisateurs se rapprocherait de ce que John Battelle appelle « *Database of Intentions* »³⁵ (la « base de données de nos intentions »).

La motivation en faveur du processus de personnalisation et d'adaptation dépend des acteurs impliqués. L'utilisateur compte sur une limitation des résultats non pertinents et espère obtenir des liens vers des informations restreintes, mais pertinentes. Du point de vue du fournisseur du moteur de recherche, la personnalisation offre la possibilité de cibler davantage la publicité, en permettant aux abonnés publicitaires de vendre des « globes personnalisés » (« *personalised eyeballs* »). Elle augmente le taux de fréquentation et réduit éventuellement le risque de clics frauduleux. Du point de vue technologique, la personnalisation constitue un compromis entre la performance au niveau du traitement et le délai de réponse. D'une part, l'augmentation des informations générées par les utilisateurs contribue à limiter le champ des recherches et, d'autre part, le volume d'informations à traiter est supérieur. Si l'on considère que la recherche audiovisuelle et la reconnaissance des formes font appel à des processus très lourds, la possibilité d'introduire une requête en langage naturel risque de ralentir la réponse.

La personnalisation peut intervenir au niveau de la recherche individuelle, comme A9, Jeeves, Google, suivre des moteurs de recherche collectifs tels qu'Eurekster ou bien se faire par domaine (historique des recherches, *tagging*, partage des notes et des journaux, etc.). Pratiquement tous les grands moteurs de recherche utilisent plusieurs méthodes pour optimiser les recherches. La plus fréquente consiste à utiliser des cookies assortis d'identifiants uniques reliés au navigateur. Cela permet au serveur de garder en mémoire un journal comportant les mots-clefs, l'adresse IP, la langue, la date et l'heure, le type de navigateur, l'URL des pages demandées, les réacheminements, etc. D'autres données pertinentes peuvent être extraites à partir de sources diverses, notamment la recherche sur

31) <http://www.clusty.com/> (consulté le 6 juin 2008).

32) <http://www.kartoo.com/> (consulté le 5 juin 2008).

33) <http://www.mister-wong.com/> (consulté le 5 juin 2008).

34) <http://search.chacha.com/> (consulté le 9 juin 2008).

35) Battelle, John (Sept. 2005). *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York, ISBN 1-59184-088-0.

ordinateur, la recherche à partir d'une barre d'outils ou la recherche géolocalisée. Le risque existe de voir ces informations reliées et recoupées avec des données provenant d'autres sources et applications, par exemple en combinant les données laissées dans « Recherche Google » par un utilisateur avec les informations extraites de son compte Gmail.

1.2.4. Vers une recherche à partir du contenu

La recherche AV fait souvent référence de façon spécifique à des techniques mieux connues sous le nom de recherche à partir du contenu. Ces moteurs peuvent rechercher des contenus audiovisuels en utilisant essentiellement la technologie de reconnaissance des formes ou vocale, pour trouver des éléments similaires parmi différents fichiers images ou audio³⁶.

Les techniques de reconnaissance vocale ou des formes permettent de prendre en compte les caractéristiques propres de l'image (par exemple sa forme ou sa couleur), ou du fichier audio. À l'avenir, ces moteurs pourront rechercher et reconnaître l'expression « montagne rouge » dans une chanson ou repérer si un fichier image ou vidéo comporte une « montagne rouge », même en l'absence d'étiquette textuelle annexée au fichier fournissant l'indication correspondante³⁷.

Les méthodes de recherche à partir du contenu sont donc indispensables lorsque les annotations textuelles sont incomplètes ou inexistantes. Elles peuvent également renforcer la précision de la recherche même lorsque des annotations textuelles existent. Le défi consiste à trouver un moyen d'améliorer la recherche multimédia en utilisant les méthodes de recherche à partir du contenu. À cet égard, la performance du processus de recherche et la qualité des résultats dépendent du type de contenu, comme le constatent Lew et autres dans une étude globale³⁸.

De nombreux concepts (c'est-à-dire algorithmes) et systèmes ont été développés pour prendre en charge la structuration, l'indexation et la recherche automatiques de *contenus vocaux et audio*.

Dans les années 90, le système de recherche d'émissions radiodiffusées « THISL broadcast news retrieval » fut sans doute l'une des premières applications pour la recherche de documents vocaux³⁹. Le discours était converti en texte à l'aide d'une vaste base de données lexicale en discours continu. Un code temporel était attribué à chaque mot des séquences lexicales transcrites automatiquement. Ce système permettait une recherche à partir d'un dispositif de recherche textuelle standard. Depuis, des progrès importants ont été réalisés, comme en atteste les exercices d'évaluation du National Institute of Science and Technology (NIST) des États-Unis. Le NIST effectue régulièrement des tests de performances dans le cadre de campagnes d'évaluation de la qualité de la recherche de documents parlés. En comparant les performances d'indexation des journaux d'actualité radiodiffusés lors des conférences successives consacrées à la recherche textuelle (TREC, 1997-2000), il est apparu clairement que le défi croissant portait sur l'amélioration significative de la qualité des systèmes. Cela s'explique, entre autres, par le fait que la recherche de discours enregistrés est très complexe : la performance globale dépend d'un nombre considérable de facteurs, notamment du nombre de locuteurs, du bruit de fond, du type de discours (émission radiodiffusée ou conversation), etc. et chacun de ces facteurs constitue un véritable défi à part entière.

La difficulté est la même dans le domaine de la recherche musicale. L'un des premiers systèmes d'indexation et de recherche de *musique* était basé sur les caractéristiques élémentaires du traitement audio⁴⁰. Il s'agissait principalement de retombées des systèmes inventés pour les normes audio MPEG7 et Audio-ID par différents groupes. La technologie Audio-ID génère une empreinte à partir d'un

36) La technologie de reconnaissance des formes et de la parole peut être un outil pertinent pour identifier les contenus et empêcher la diffusion de documents protégés par le droit d'auteur. Voir Anick Jesdanun, *Myspace Launches Pilot To Filter Copyright Video Clips, Using System From Audible Magic, Associated Press Newswires*, 12 février 2007.

37) Voir Dr. Fuhui Long, Dr. Hongjiang Zhang et Prof. David Dagan Feng, *Fundamentals of Content-Based Image Retrieval*, sur http://research.microsoft.com/asia/dload_files/group/mcomputing/2003P/ch01_Long_v40-proof.pdf (consulté le 6 juin 2008).

38) Lew, M. S., Sebe, N., Djeraba, C., et Jain, R. 2006. *Content-based multimedia information retrieval: State of the art and challenges*. ACM Trans. Multimedia Comput. Commun. Appl. 2, 1 (février 2006), 1-19. DOI= <http://doi.acm.org/10.1145/1126004.1126005>

39) D. Abberley, D. Kirby, S. Renals and T. Robinson, *The THISL broadcast news retrieval system, Proceedings of ESCA ETRW Workshop on Accessing Information in Spoken Audio*, Cambridge (UK), avril 1999.

40) Foote, J. 1999. *An Overview of Audio Information Retrieval*. *ACM Multimedia Systems* 7(1), 42-51.

segment musical et fournit des algorithmes de couplage rapide pour retrouver cette empreinte dans une vaste base d'archives prétraitées. Dernièrement, la recherche de musique s'est réorientée vers la classification en fonction du genre et de l'ambiance.

Pour la *recherche de vidéos*, au milieu des années 90, les travaux visaient à trouver un moyen de détecter le contour des objets et l'écart entre les histogrammes chromatiques correspondant à deux images successives d'une vidéo⁴¹. Depuis, des progrès ont été réalisés au niveau de l'automatisation du processus de repérage des contours sur un mode objectif et une méthode a vu le jour consistant à utiliser le mouvement au sein de la vidéo pour segmenter les plans ; parallèlement, la classification sémantique des séquences vidéo a été développée sur la base de catégories telles que zoom avant, zoom arrière, panorama, etc.⁴²

Le secteur des moteurs de recherche est actuellement en pleine expansion et plusieurs versions Bêta sont disponibles pour les différentes méthodes de recherche appliquées aux informations visuelles et sonores.

Tiltomo⁴³ et Riya⁴⁴ fournissent des outils évolués de recherche d'images à partir du contenu, capable d'identifier des concordances à partir d'index basés sur les couleurs et les formes de l'image demandée. Pixsy⁴⁵ collecte des contenus visuels à partir de milliers de fournisseurs du Web, et permet de rechercher ces images et vidéos sur la base de leurs caractéristiques visuelles.

TVEyes⁴⁶ et Audioclippping⁴⁷ ont créé un index lexical sonore à l'aide d'une technologie sophistiquée de reconnaissance vocale, permettant aux utilisateurs de rechercher des émissions de radio, télévision ou des podcasts à partir de mots-clefs⁴⁸. Blinkx⁴⁹ et Podzinger⁵⁰ utilisent l'analyse visuelle et la reconnaissance vocale pour une indexation optimisée des contenus multimédia enrichis sous un format audio ou vidéo.

Néanmoins, dans un avenir proche, le scénario le plus probable sera la convergence et la combinaison de la recherche textuelle avec les technologies de recherche indexant les informations audio et visuelles⁵¹. Pixlogic⁵², par exemple, offre la possibilité de rechercher non seulement les métadonnées d'une image spécifique, mais aussi des fragments d'image qui peuvent faire l'objet d'une recherche.

Le tableau ci-après récapitule les principales méthodes technologiques actuelles et leurs défis pour l'analyse multimédia, en lien avec les applications d'avenir actuelles et futures. Le tableau est organisé en fonction des différents formats de médias, notamment l'analyse textuelle, sonore et visuelle. Le domaine audio fait la distinction entre la parole et la musique, compte tenu des différentes gammes d'application et des défis correspondants. Certains défis majeurs de la recherche se retrouvent dans différents domaines, tels que la question précédemment évoquée de la sémantique et de la profusion d'informations.

41) Flickner, M. Sawhney, H. Niblack, W. Ashley, J. Qian Huang Dom, B. Gorkani, M. Hafner, J. Lee, D. Petkovic, D. Steele, D. Yanker, P. 1995. *Query by image and video content: the QBIC system*, *IEEE Computer*, septembre, 23-32.

42) Lienhart, R. 2001. *Reliable Transition Detection in Videos: A Survey and Practitioner's Guide*. *International Journal of Image and Graphics* 1(3), 469-486.

43) <http://www.tiltomo.com/> (consulté le 6 juin 2008).

44) <http://www.riya.com/> (consulté le 6 juin 2008).

45) <http://www.pixsy.com/> (consulté le 6 juin 2008).

46) <http://www.tveyes.com/> (consulté le 6 juin 2008); TVEyes propose un service appelé Podscope (<http://www.podscope.com/>) (consulté le 6 juin 2008) qui permet aux utilisateurs de rechercher le contenu de podcasts placés sur le Web.

47) <http://www.audioclippping.de/> (consulté le 6 juin 2008).

48) Voir Gary Price, *Searching Television News*, *SearchEngineWatch*, 6 février 2006, sur <http://searchenginewatch.com/showPage.html?page=3582981> (consulté le 6 juin 2008).

49) <http://www.blinkx.com/> (consulté le 6 juin 2008).

50) <http://www.podzinger.com/> (consulté le 6 juin 2008).

51) Voir Brendan Borrell, *Video Searching by Sight and Script*, *Technology Review*, 11 octobre 2006, sur http://www.technologyreview.com/read_article.aspx?ch=specialsections&sc=personal&id=17604 (consulté le 6 juin 2008).

52) <http://www.pixlogic.com/> (consulté le 6 juin 2008).

Tableau 1 : Récapitulatif des méthodes, des applications et des défis liés aux technologies de recherche à partir du contenu en fonction du type de format.

Texte	Parole	Musique	Images	Vidéo
Méthode technologique				
<ul style="list-style-type: none"> - Reconnaissance d'une entité nommée, SVM, PLSI raisonnement sémantique bayésien - Augmentation des légendes 	<ul style="list-style-type: none"> - Reconnaissance vocale : recherche de documents parlés, indexation de sous-séquences - Segmentation du discours : regroupement et reconnaissance de locuteurs transcription discours - vidéos 	<ul style="list-style-type: none"> Segmentation musicale : platitude spectrale algorithmes génétiques, recherche et recommandation de musique 	<ul style="list-style-type: none"> - Traitement d'images brutes (histogrammes, formes, textures, documents MPEG7, SIFT), mesure de similarité des images - Retour de pertinence 	<ul style="list-style-type: none"> - Détection des plans, création de séquences-clefs - Suivi d'objet à partir des paramètres de mouvement, reconnaissance de sous-titres codés, etc. - Détection et reconnaissance d'objet - Annotation et résumé de vidéos - Détection d'évènement vidéo
Applications				
<ul style="list-style-type: none"> - Classification des actualités et des archives dans les entreprises - Filtrage d'e-mail - Moteurs de recherche textuels, analyse sémantique d'annotations (automatiques) multimédia 	<ul style="list-style-type: none"> - Indexation de journaux d'actualité/archives de radiodiffusion - Recherche de podcasts/vidéocasts (Potzinger, Blinkx) archives audio (données parlementaires, archives historiques) 	<ul style="list-style-type: none"> - Indexation de musicothèques - Recherche en fredonnant une mélodie - identification audio et musicale, moteurs de recommandation 	<ul style="list-style-type: none"> - Recherche à partir du contenu dans des iconothèques - Reconnaissance d'objet, reconnaissance faciale (sécurité, photothèques), annotation automatique d'iconothèques par mots-clefs et descripteur textuel 	<ul style="list-style-type: none"> - Indexation de documents de radiodiffusion, observation des médias, indexation de vidéos, moteurs de recommandation, création d'empreintes de vidéos, détection de logos, sécurité, etc. - Vidéos 3D
Défis				
<ul style="list-style-type: none"> - Sémantique, conception d'ontologies (par ex. enrichissement) 	<ul style="list-style-type: none"> - Variabilité des contenus (par ex. bruit de fond), contexte lié à la discipline et la langue, évolutivité de la méthode des sous-séquences 	<ul style="list-style-type: none"> - Classification par genre - Reconnaissance d'instruments polyphoniques, - Analyse subjective 	<ul style="list-style-type: none"> - Fossé sémantique, segmentation des images, fossé sensoriel 	<ul style="list-style-type: none"> - Détection de concepts complexes, de milliers d'objets différents. Segmentation en unités plus sémantiques (scènes complexes), multimodalité, fusion

Source : adaptation de Sebe/Koehler⁵³

53) N. Sebe, J. Koehler, *State of the art in audio-visual content indexing and retrieval technologies* dans *Chorus Report 2.1 State of the Art on Multimedia Search Engines*, p.29 et suivantes, disponible sur : http://www.ist-chorus.org/documents/D21_final.pdf (consulté le 10 juin 2008).

2. Considérations économiques

Du fait de l'explosion des données dans le domaine numérique, la recherche est devenue une fonctionnalité complexe. L'abondance des données est à l'origine du développement des moteurs de recherche, mais le fait que la recherche soit tributaire du contexte introduit une distinction entre les applications et leurs marchés. Les moteurs de recherche bien connus sur le Web proposent aux internautes un service de recherche de contenus publiquement accessibles sur l'Internet. Leur modèle commercial est essentiellement soutenu par la publicité, ce qui est très différent, par exemple, des moteurs opérant dans le cadre de la recherche en entreprise. Ces derniers permettent d'effectuer des recherches le plus souvent dans des bases de données multimédia propriétaires fermées (bibliothèques nationales, par exemple). Les prestataires de systèmes fournissent un outil de recherche spécifiquement conçu en fonction des besoins de l'entreprise cliente et adapté à ses exigences en matière de sécurité, de droits d'accès, de sécurisation des contenus, etc.

Si les outils de recherche sont confrontés aux mêmes défis technologiques, la nature de leurs applications est différente, et il en va de même pour leurs marchés. Par conséquent, les quatre principaux domaines en jeu, à savoir la recherche sur le Web, la recherche en entreprise, la recherche *peer-to-peer* et la recherche mobile, seront abordés de façon spécifique dans les chapitres suivants.

2.1. Recherche sur le Web

En dépit du fait que près d'une centaine de moteurs de recherche sont opérationnels, l'essentiel des recherches est effectué par un très petit nombre de prestataires. Selon le cabinet-conseil Nielsen/Netratings, les trois principaux opérateurs contrôlent plus de 80 % du marché. En particulier, 53,6 % des recherches en ligne effectuées aux États-Unis en août 2007 ont été effectuées par Google, 19,9 % par Yahoo!, 12,9 % par MSN, 5,6 % par AOL, 1,7 % par Ask 1,7 % et 6,3 % par le reste des opérateurs. Ces recherches englobent les recherches locales, les recherches d'images, les recherches d'actualité, les recherches d'achat et autres types de recherche verticale. Plus de 5,6 milliards de recherche ont été effectuées sur la seule période de ce mois (août 2007)⁵⁴.

Les trois principaux opérateurs se classent nettement en tête. Conformément à comScore Networks, en décembre 2006, les sites Google détenaient 47,3 % du marché US des recherches, Yahoo! 28,5 % et Microsoft 10,5 %. Les américains ont effectué 6,7 milliards de recherches en décembre 2006. Par rapport au même mois de l'année précédente, cela représente une progression annuelle de 30 % du volume des demandes de recherche. Ce taux de progression est considérable et explique pourquoi la publicité en ligne sur les moteurs de recherche est considérée comme un marché de croissance très prometteur.

Les internautes européens utilisent les moteurs de recherche de façon aussi intensive que leurs semblables d'Outre-Atlantique. La fréquence d'utilisation des moteurs de recherche explique pourquoi ils figurent parmi les sites les plus visités d'Internet et connaissent un trafic considérable. Google est le moteur de recherche le plus fréquenté dans pratiquement tous les pays de l'Union européenne. En juin 2007, par exemple, Google a attiré 88,8 % de la population en ligne du Royaume-Uni, 69,5 % en France et 69 % en Allemagne. Le niveau d'audience de Google est largement supérieur aux sites de Microsoft (83,3 % au Royaume-Uni, 62,3 % en France, 54 % en Allemagne) de Yahoo! (65,9 % au Royaume-Uni, 39,6 % en France et 36 % en Allemagne), selon la société d'audimétrie Internet comScore⁵⁵. La consolidation du marché des moteurs de recherche apparaît de façon manifeste lorsqu'on observe l'évolution des résultats sur des périodes plus longues.

Les figures 4 et 5 montrent l'évolution des parts respectives pour l'Allemagne et la France. L'évolution de l'Allemagne et de la France est similaire aux autres États membres de l'UE. On constate en particulier qu'une poignée de fournisseurs se partagent plus 90 % des parts de marché et que la part de Google est largement supérieure à celle de ses concurrents.

54) Voir <http://www.nielsen-netratings.com/>

55) Communiqués de presse de comScore disponibles sur <http://www.comscore.com/>

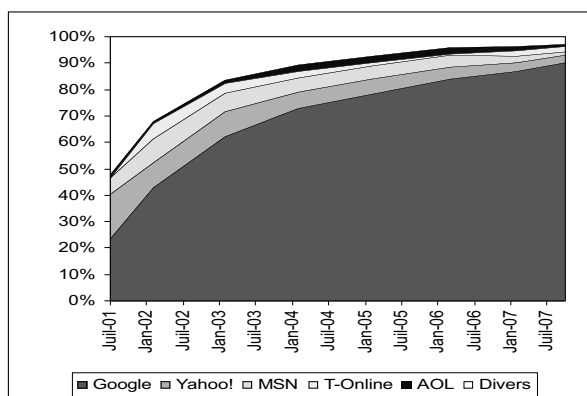


Figure 4 : Évolution de la fréquentation des moteurs de recherche en Allemagne sur la période 2001-2007.

Source : WebBarometer,⁵⁶ H Speck⁵⁷ et calculs personnels.

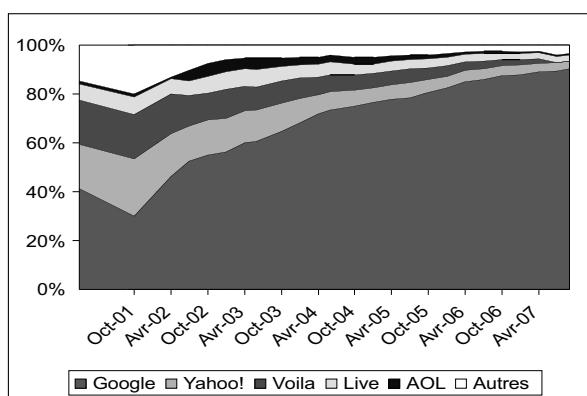


Figure 5 : Évolution de la fréquentation des moteurs de recherche en France sur la période octobre 2001 - septembre 2007.

Source : Baromètre Secrets2Moteurs⁵⁸ et calculs personnels.

Ces données mettent en lumière, d'une part, que le marché des moteurs de recherche est fortement concentré et, d'autre part, que ces derniers font désormais partie intégrante de notre quotidien. En moyenne, un internaute allemand utilise Google plus de quarante fois par mois⁵⁹ et les trois-quarts des internautes accèdent aux offres d'Internet par le biais des moteurs de recherche⁶⁰. Même si le niveau d'audience des prestataires de moteurs de recherche varie d'un pays à l'autre, le comportement des utilisateurs est similaire dans la plupart des pays occidentaux.

Les cabinets-conseil qui calculent les parts de marché tels que Nielsen/NetRatings, Compete, Hitwise ou comScore, étudient le comportement des internautes en matière de recherche en installant des compteurs en temps réel sur leur ordinateur (Nielsen revendique un panel de 500 000 internautes au niveau mondial). Les parts de marché établies par ces cabinets-conseil peuvent varier, dans une certaine mesure, de celles établies par chacun des moteurs de recherche, du fait qu'ils emploient des outils de mesure différents et que la précision des données n'est pas clairement définie. Ce phénomène permet d'expliquer en partie pourquoi les données d'audimétrie de comScore pour l'Allemagne et la France diffèrent des résultats enregistrés par l'allemand WebHits et le français Secrets2Moteurs.fr.

56) <http://webhits.de/>

57) Hendrik Speck. *Search.Market*. 10th InetBib Conference. 9-11 avril 2008, Würzburg (Allemagne) disponible sur : <http://www.egs.edu/faculty/speck/files/presentation2008inetbibsearchenginemarkets.pdf> (consulté le 5 juin 2008).

58) <http://www.secrets2moteurs.fr/>

59) Données comScore sur l'Allemagne, juin 2007.

60) *Internetverbreitung in Deutschland: Potenzial vorerst ausgeschöpft?* Birgit van Eimeren, Heinz Gerhards et Beaste Frees, Media Perspektiven, Vol 8, p. 350 - 370.

La concentration du marché des moteurs de recherche semble être une tendance générale aux États-Unis et dans la plupart des États membres de l'UE. Un certain nombre d'éléments permet d'expliquer pourquoi Google est plus dominant en Europe qu'aux États-Unis, notamment un meilleur impact de l'image de marque, des stratégies de marketing nationales, une meilleure adaptation à la taille du marché, une meilleure adaptation technologique au langage, l'absence de puissants moteurs de recherche nationaux, etc.

La Russie est un exemple intéressant, même si elle ne fait pas partie de l'Union européenne. Dans ce pays, Google n'arrive qu'en troisième position après Yandex⁶¹ et Rambler. Yandex se targue d'avoir une technologie supérieure, qui lui permet de mieux maîtriser les déclinaisons et la conjugaison de la langue russe que les autres moteurs de recherche. D'autres moteurs de recherche slaves tels que le tchèque Morfeo⁶² ou le polonais NetSprint⁶³ annoncent une technologie plus performante. Néanmoins, on ne peut déterminer avec clarté dans quelle mesure le taux élevé de parts de marché de Yandex en Russie (plus de 55 %) peut être attribué aux meilleures performances linguistiques de ce moteur, puisque le même fournisseur ne détient que 16 % des parts de marché en Ukraine, alors que, linguistiquement parlant, la langue russe et l'ukrainien, sont très proches.

L'un des facteurs qui a contribué à propulser Google en tête du marché est la rapidité avec laquelle il a réussi à intégrer des services innovants. Un grand nombre d'entre eux ont été proposés au public dès la phase de développement (versions bêta), plutôt que d'attendre de pouvoir offrir aux utilisateurs des services finalisés. Cette implication des utilisateurs à un stade précoce de développement fait partie de la culture d'apprentissage par la pratique propre à l'entreprise. Google a profité de l'usage prédominant de l'anglais sur Internet pour tester ses services et applications auprès d'un large public anglo-saxon, avant de les adapter et de les lancer dans d'autres contextes culturels.

Coût d'exploitation

Le coût d'exploitation pour fournir un outil de recherche sur le Web est énorme.

Schématiquement, ce coût se répartit en deux volets principaux : d'une part, le coût de l'exploration et de l'indexation des contenus, qui est pratiquement proportionnel au volume des contenus. Les données numériques se sont considérablement développées et rien n'indique que le nombre de sites et de données ne va pas continuer à progresser de façon démesurée à l'avenir. Par conséquent, les coûts liés à la collecte et l'indexation des contenus vont continuer d'évoluer à la hausse. L'exploration est une tâche qui exige des ressources considérables, mais ne constitue pas un facteur discriminant significatif parmi les fournisseurs de moteurs de recherche qui ont leur propre système d'indexation. Au 1^{er} juin, par exemple, Exalead, un « modeste » moteur de recherche, avait exploré plus de 8 064 millions de sites, ce qui correspond au même nombre d'exploration que les gros moteurs de recherche.

D'autre part, le second volet des coûts d'exploitation est lié au coût du traitement des requêtes des internautes. Les internautes actuels sont très exigeants et attendent une réponse quasi-immédiate à leurs demandes. La capacité de traitement requise pour répondre aux demandes est proportionnelle au trafic, c'est-à-dire au parc d'utilisateurs. Il en va de même pour les coûts. Pour les gros moteurs de recherche ayant un taux de fréquentation élevé, ces coûts sont énormes. Dans une entreprise bien conçue et bien gérée, les recettes devraient également être proportionnelles au trafic et au parc d'utilisateurs. Par conséquent, la fréquentation d'un moteur de recherche pourrait être progressivement compensée par l'équilibre entre les dépenses et les recettes. Néanmoins, ce type de croissance organique exige d'être soutenue par d'autres facteurs tels que l'exploitation d'une technologie de recherche de pointe, une bonne image de marque, des campagnes publicitaires bien conçues, un équilibre financier, etc., tous ces éléments constituant autant d'obstacles pour qu'un nouveau venu soit en mesure de rattraper les leaders du marché. Cette barrière à l'entrée du marché va probablement s'accroître au fur et à mesure que la recherche se déplace du texte vers les contenus audiovisuels, ce qui implique à la fois une technologie plus complexe et une augmentation continue des coûts d'exploitation.

61) <http://www.yandex.ru/>

62) <http://morfeo.centrum.cz/>

63) <http://www.netsprint.pl/serwis/>

Les développements technologiques peuvent contribuer à la réduction des coûts d'exploitation. Cela suppose des algorithmes plus rapides, des architectures serveurs plus performantes, du matériel plus économique et des processeurs moins gourmands en énergie. En fait, la réduction de la consommation d'énergie des parcs de serveurs constitue un véritable défi. Les parcs de serveurs évolués ont une telle consommation d'énergie que le prix de l'électricité et l'accès aux ressources de refroidissement (par ex. rivières ou lacs) figurent parmi les principaux critères de location.

La recherche, pivot central entre spécialisation et intégration des services

Parallèlement à l'essor de la recherche comme un outil technologique central pour les citoyens et les entreprises, les exigences des utilisateurs concernant les performances des moteurs et la pertinence des résultats n'ont cessé d'évoluer. Le besoin de fournir des résultats de plus en plus pertinents explique pour une part la concentration des moteurs de recherche généralistes et l'émergence de moteurs thématiques. Dans le premier cas, seuls quelques fournisseurs peuvent supporter la pression due à la hausse des coûts d'exploitation. Ces coûts englobent la création de parcs de serveurs, leur maintenance, les frais d'exploitation, les coûts de développement pour améliorer leur performance, ainsi que la création de nouveaux services annexes. Par conséquent, seuls quelques moteurs de recherche généralistes sont en mesure de prospérer en opérant au niveau international. Si l'on considère que les contenus audiovisuels requièrent une capacité de traitement plus puissante que pour le texte, il est peu probable que la barrière d'accès au marché s'atténue pour les nouveaux opérateurs de nouveaux moteurs de recherche audiovisuels. On assistera probablement à une consolidation autour d'un petit nombre de moteurs de recherche généralistes. La spécialisation sur un créneau spécifique peut également offrir une alternative. C'est le cas des moteurs thématiques qui fournissent une recherche de contenus pertinents dans des domaines spécifiques ; le marché des créneaux spécialisés est plus facilement accessible que celui des moteurs de recherche généralistes.

Les moteurs de recherche sont de plus en plus fréquemment intégrés dans d'autres applications et services. Les principaux moteurs de recherche proposent des plateformes de services gratuits tels qu'e-mail, forum, chat, messagerie instantanée, services vocaux, etc. Nombre de ces services ne sont pas rentables en tant que tels, mais au niveau global, il est intéressant de conserver les utilisateurs dans la « mouvance » de l'opérateur du moteur de recherche. Étant donné que les fournisseurs de moteurs de recherche ont tout intérêt à « fidéliser » leurs utilisateurs, ils s'efforcent de personnaliser les services. La personnalisation leur permet d'être en mesure d'anticiper les demandes et de promouvoir un mode de fonctionnement proactif des moteurs de recherche. Au lieu de simplement « extraire » les informations, ils pourraient également les « semer » et, par la même occasion, mieux cibler la publicité.

2.2. Les systèmes de recherche en entreprise

Autrefois, les investissements d'une entreprise étaient majoritairement consacrés à la mise en place d'une infrastructure informatique efficace. Au fil du temps, les entreprises ont rassemblé les ressources et les technologies requises pour collecter, stocker et transférer les informations nécessaires à leur fonctionnement. Il reste à trouver une solution pour fournir un accès consolidé centré sur l'utilisateur pour faciliter le travail des employés tout en leur permettant d'être plus efficaces. Cette transition d'une infrastructure orientée mémoire vers une consommation des informations va de pair avec la mise en place d'un modèle davantage centré sur l'utilisateur que sur une technologie. Le développement d'un système efficace, interactif et sécurisé de présentation de contenus centré sur l'utilisateur est une tâche complexe, car il faut prendre en compte différents systèmes d'exploitation, différents formats de fichiers, différents schémas, etc. Un exemple de solution de recherche en entreprise nous est fourni par les outils intelligents des prestataires de services de pages jaunes (et blanches) de l'annuaire téléphonique. Ces outils de recherche spécifiques extraient tout d'abord les informations à partir des bases de données des entreprises, puis les recoupent avec d'autres sources. Un processus d'analyse et de comparaison permet de repérer les données manquantes ou incompatibles. Le secteur de la santé constitue un autre exemple. Les dossiers, images, radios, analyses, scanners et autres informations médicales pertinentes sont de plus en plus générées sous une forme numérique. Ce secteur a besoin d'outils automatiques de recherche, d'analyse et d'interprétation de ce type d'informations.

Par conséquent, les solutions de recherche sur mesure pour le commerce et l'industrie constituent un secteur d'activité en voie de développement. Les entreprises peuvent créer de la plus-value en identifiant et en diffusant des contenus spécifiques au sein de leur organisation, avec la possibilité pour

les utilisateurs agréés de les indexer, de les rechercher et de les consulter. Selon un rapport du cabinet-conseil IDC, le marché mondial de la recherche en entreprise et des logiciels de recherche correspondants se chiffrait à 976 millions de dollars US (USD) en 2005. Ce chiffre représente une progression de 32 % par rapport à l'année précédente. Ce secteur est manifestement plus restreint que le marché de la publicité sur la recherche Web dont nous avons parlé précédemment. Les trois grands acteurs Google, Yahoo! et Microsoft mènent quelques activités dans ce domaine, mais leurs revenus générés par les licences technologiques sont modestes par rapport à leur activité principale. Contrairement au marché de la recherche sur le Web, le marché des solutions d'entreprises est fragmenté et il existe un certain nombre de sociétés très spécialisées sur ce marché. Il existe également plusieurs entreprises européennes, dont les plus réputées sont, entre autres, la société norvégienne FAST (qui vient d'être reprise par Microsoft), Autonomy au Royaume-Uni ou Expert System SpA en Italie. Ces entreprises vendent des outils de gestion des connaissances où la fonction de recherche constitue un module de plus en plus important. Grâce à ces outils, les entreprises entendent rechercher les informations de façon efficace et exploiter la signification attachée à toute sorte de communication d'entreprise, qu'il s'agisse de documents, d'e-mails, de saisies dans des bases de données relationnelles, etc.

Actuellement, le marché des outils de « gestion des connaissances » est très différent du marché des moteurs de recherche Web. Néanmoins, l'avenir de l'Internet est davantage audiovisuel que textuel.

Plus les moteurs de recherche Web seront confrontés à des contenus audiovisuels, plus les intérêts technologiques liés à la nécessité de développer des solutions de recherche conceptuelle, de classification des documents, de fouille de textes, d'analyse et de corrélation des informations vont converger. Au cours de cette évolution, les moteurs de recherche Web actuels seront sans doute amenés à pénétrer davantage le marché de la « gestion des connaissances ». La récente reprise de FAST par Microsoft peut être interprétée comme un indicateur de cette tendance⁶⁴.

2.3. La recherche mobile

La recherche mobile désigne les services de recherche d'information accessibles via des appareils mobiles tels que téléphone portable ou PDA. Les opérateurs télécoms européens proposent d'ores et déjà des options de recherche dans le cadre de leurs services 2G (GSM), 2,5G (GPRS) et 3G (UMTS), sur la base de la technologie fournie par des tiers telles que Google ou FAST, et de nombreux portails de recherche offrent une interface dédiée pour les services mobiles, tels que MetaGer.

Bien que les perspectives concernant les modèles commerciaux ne soient pas encore consolidées, le marché de la recherche mobile sera probablement très différent du marché des moteurs de recherche Web. L'environnement technologique (écrans de taille réduite, largeur de bande limitée), le nombre restreint de contenus adaptés aux appareils mobiles, le rôle des acteurs du marché (les opérateurs télécoms fournissant un accès Internet sur les portables jouent un rôle plus important que les fournisseurs d'accès Internet via l'ordinateur), le comportement des utilisateurs (type de recherche effectué en situation nomade), tous ces facteurs contribueront sans doute à l'émergence d'un modèle commercial de moteur de recherche différent. Les marchés cloisonnés semblent être le modèle dominant, mais le marché peut évoluer pour s'ouvrir davantage. La possibilité d'une tarification au forfait est en cours de discussion et les restrictions au niveau de la largeur de bande pourraient imposer un paiement en fonction du volume d'octets téléchargés. Cela introduirait une différence non seulement pour les téléchargements utilisant un débit élevé, tels que les vidéos (par ex. il pourrait y avoir un tarif par vidéo en fonction de la résolution), mais aussi pour les services de géolocalisation qui sont considérés comme très prometteurs et nous permettront de trouver le restaurant le plus proche par une simple saisie vocale ou clavier de la question sur le téléphone portable.

Même si ce marché n'en est qu'à ses débuts et reste fragmenté, les tendances sont extrêmement prometteuses et étroitement liées au nombre d'abonnés aux connexions mobiles à très haut débit. Si les chiffres restent minimes en termes absolus, ils présentent néanmoins, un taux de croissance impressionnant. La GSM Association annonçaient 32 millions de connexions mobiles à haut débit en mars 2008, soit 3 millions de plus qu'en mars 2007⁶⁵! Qui plus est, l'Internet et la recherche mobile

64) <http://www.01net.com/editorial/368946/microsoft-s-achete-la-place-de-numero-un-de-la-recherche-en-entreprise/>

65) www.gsmworld.com

offrent un marché en pleine expansion. Son principal attrait consiste en la fourniture d'informations locales, telles que la météo, les plans ou les itinéraires. On estime qu'en 2008, il y aura 2,5 milliards d'utilisateurs de portables dans le monde, dont environ 40 % disposent de la technologie 2,5G et 10 % du 3G. Étant donné que le nombre d'abonnés de la téléphonie mobile est toujours en hausse au niveau mondial (surtout dans les pays à forte population et en voie de développement, où le taux de pénétration est réduit) et que la part des portables 2,5G ou plus va augmenter, le nombre d'abonnés va également s'accroître. Actuellement, 489 millions de personnes ont accès à l'Internet mobile (pas nécessairement à haut débit) et ce chiffre pourrait doubler d'ici 2011. Étant donné que la grande majorité des internautes mobiles aura également recours à la recherche d'informations, cela constitue une clientèle potentielle pour les fournisseurs de moteurs de recherche. En fait, eMarketer estime que le marché publicitaire de la recherche mobile se chiffre à 221 millions USD en 2008 et atteindra 2 361 millions USD d'ici 2011⁶⁶.

Néanmoins, pour que ces prévisions se réalisent, il faudra qu'un certain nombre d'obstacles, liés aux exigences spécifiques concernant l'interaction des usagers, la recherche et l'affichage des informations, soit surmonté. Par ailleurs, l'adaptation ou la création de contenus compatibles avec les appareils mobiles reste un problème. La recherche de contenus ajoutera de la plus-value lorsque les contenus seront adaptés à l'utilisateur et associés à d'autres technologies, par exemple les services de géolocalisation. Cette démarche permettra de personnaliser le processus de recherche des utilisateurs.

Expansion du processus de recherche

Ces outils sont conçus pour rechercher des informations sur des plateformes spécifiques. La recherche actuelle sur le Web est majoritairement effectuée sur des ordinateurs personnels ou des portables qui ont accès à l'un des quelque 70 moteurs de recherche en service dans le monde entier. Le défi consiste, entre autres, à élargir les capacités de recherche à d'autres plateformes, notamment aux appareils mobiles. Les possibilités technologiques relatives aux plateformes autres que les ordinateurs n'en sont qu'à leurs débuts. Par exemple, pour la recherche mobile, les appareils sont limités au niveau technique et les modèles commerciaux ne sont pas encore consolidés. Du point de vue de l'utilisateur, l'idéal serait de disposer d'une interopérabilité intégrale, puisque leur seul intérêt est de rechercher des contenus pertinents, quelle que soit la technologie.

3. Synthèse

Les outils de recherche actuels (le plus souvent textuels) constituent des technologies décisives pour la société de l'information. Du fait de l'explosion des données audiovisuelles, les futurs moteurs de recherche AV vont jouer un rôle croissant dans la société. La fourniture de résultats pertinents dans la recherche audiovisuelle est beaucoup plus complexe que dans la recherche textuelle, et l'évolution de la recherche AV dépendra des futures innovations et de l'amélioration des concepts existants. Ainsi, la recherche de matériel audiovisuel pertinent bénéficiera des nouveaux procédés de recherche à partir du contenu (audio, vidéo, images), tout en étant largement tributaire des concepts de métadonnées (textuelles). Les défis technologiques concernent aussi bien la science fondamentale que les tâches de développement. Les tâches de recherche fondamentale portent, entre autres, sur les méthodes sémantiques de recherche ou de nouveaux concepts technologiques permettant de maîtriser la création et le flux d'énormes quantités de données. Les concepts de recherche interactive, les systèmes de retour de pertinence, l'analyse multimodale ou les algorithmes de recherche évolués sont autant d'éléments à résoudre pour pouvoir améliorer les applications actuelles.

Les considérations technologiques sont directement liées à la situation économique. La gestion du traitement, du stockage et de la circulation d'énormes volumes de données audiovisuelles constitue un défi technologique qui exigera la mise en place de nouvelles architectures informatiques et de solutions de recherche répartie. Parallèlement, il y a d'importantes implications financières qui pourraient avoir une influence considérable sur l'accès de nouveaux opérateurs sur le marché de la

66) www.eMarketer.com

recherche audiovisuelle. Le déploiement de la technologie de recherche AV aura donc probablement pour effet de renforcer un certain nombre de tendances technico-économiques actuelles du marché de la recherche sur le Web, notamment la concentration des moteurs de recherche généralistes sur le Web ou l'émergence de réseaux thématiques de recherche. À l'heure actuelle, le marché de la recherche se différencie clairement de la recherche sur le Web, en entreprise ou mobile, en raison de la nature du secteur, de la structuration de la clientèle et des modèles commerciaux. À l'avenir, ces différences pourraient s'atténuer. Les fournisseurs de moteurs de recherche sur le Web commencent à racheter des entreprises qui proposent des outils de recherche en entreprise, et des modèles de marchés non-cloisonnés pourraient émerger dans le secteur de la recherche mobile.

L'avenir de la navigation dans les contenus

*Janet Greco**

Broadcast Projects, Londres

Introduction

A une époque où les consommateurs sont confrontés de façon croissante à une offre pléthorique de contenus (audiovisuels et autres) rivalisant pour capter leur attention, comment peuvent-ils être sûrs de ne pas passer à côté du contenu qui correspond vraiment à leurs intérêts ? Face à un scénario de plus en plus complexe, les mécanismes de navigation dans les contenus, tels que les guides de programmes électroniques et les moteurs de recherche, jouent un rôle clé.

1. Les guides de programmes électroniques à la croisée des chemins

Autrefois, pour savoir ce qui passait à la télévision, il suffisait tout simplement de jeter un coup d'œil rapide à la poignée de chaînes répertoriées dans le journal ou le programme télévisé hebdomadaire local. Les guides des programmes au format papier fournissaient le titre et le détail des émissions à venir sous une présentation en colonnes, facile à lire, organisée de façon chronologique. Des commentaires rédactionnels supplémentaires, accompagnés d'une photo, signalaient pour chaque jour les émissions dignes d'intérêt. Le nombre de chaînes commençant à augmenter, des listes sous forme de « grille » ont fait leur apparition, toujours sur support papier ; vers le milieu des années 80, elles sont devenues le format privilégié, avant de finir par s'imposer. Depuis l'introduction des guides de programmes électroniques (EPG) pour les systèmes de télévision numérique au milieu des années 1990, ce format en « grille » est demeuré l'unité de structure élémentaire permettant la navigation dans l'environnement de la télévision à péage. Au cours de la dernière décennie, les EPG sont devenus un outil essentiel pour naviguer parmi les contenus, dans un paysage télévisuel qui s'est mué en marché de masse. Mais avec un si grand nombre de chaînes et une « surface utile d'écran » insuffisante, sans aucun moyen de transmettre la grande majorité des détails utiles qui pourraient permettre une prise de décision éclairée, ces guides sous forme de grille ont atteint leurs limites.

Plus récemment, les outils Internet sont devenus un moyen courant de localiser des contenus au-delà du poste de télévision. Avec la prolifération des appareils du type TiVo et des enregistreurs vidéo personnels/numériques, nos habitudes ont changé et nous sommes aujourd'hui attirés par ce qui est « diffusé en permanence » ou « toujours disponible ». Les choix se sont multipliés, du divertissement grand public produit par des professionnels à l'avalanche de « contenus générés par les utilisateurs » grâce aux phénomènes de partage de vidéos tels que YouTube. Le défi majeur a aujourd'hui fondamen-

* Janet Greco est fondatrice et directrice consultante chez Broadcast Projects, Londres (www.broadcastprojects.com).

talement changé : comment les consommateurs vont-ils localiser le contenu correspondant à leurs goûts personnels ? La participation active et les commentaires des utilisateurs de réseaux sociaux vont-ils modifier fondamentalement la façon dont les gens partagent et reçoivent des recommandations ?

Peu de gens, sans doute, ont réfléchi réellement aux efforts que doivent déployer les intermédiaires spécialisés (en l'occurrence les « agrégateurs de contenus » télévisés et les agences de presse) pour obtenir régulièrement les programmes d'émissions auprès des télédiffuseurs et les mettre à jour. Aujourd'hui encore, ces informations sont diffusées par les divisions « marketing » des télédiffuseurs et des chaînes, sous des formats de données très variés. Il est utile de dresser un parallèle entre ces « listings télévisés » et l'import de noms et d'adresses dans un logiciel de gestion des contacts : les noms des champs doivent être compatibles et les caractéristiques des champs doivent être au bon format (on ne peut pas, par exemple, importer des lettres dans un champ « numérique », de même qu'il ne serait d'aucune utilité d'importer le nom d'une personne dans le champ « pays »). Or la plupart du temps, ce sont des documents narratifs qui sont échangés et qui ne peuvent pas être facilement importés dans une base de données. De plus, les éditeurs ont besoin d'un flux d'informations pour une multitude de chaînes, dans un format de données unique, compatible avec les caractéristiques propres à leur système. Ils ne peuvent pas gérer des formats entrants multiples, différents pour chaque chaîne.

Par conséquent, bien qu'il existe des normes internationales en matière de métadonnées, il est extrêmement difficile d'obtenir de fournisseurs multiples qu'ils rédigent, classent et vérifient des informations descriptives d'une façon homogène. Résultat, les prestataires intermédiaires d'agrégation des données qui jouent ce rôle essentiel de traitement des données, de standardisation et de vérification, demeurent des acteurs importants dans la diffusion de ce type d'informations aux éditeurs professionnels (papier et en ligne).

Le principal défaut au sein de ce processus est l'absence de lien direct (sauf dans de rares cas) entre le prestataire intermédiaire et les centres de diffusion et de transmission des chaînes (« *play-out centres* »). Par conséquent, les informations diffusées par les agrégateurs de données sont naturellement déconnectées de la télédiffusion en elle-même, ce qui peut entraîner des erreurs. Pourtant, les fonctions d'harmonisation des données qu'ils assurent leur garantissent de conserver un rôle important à l'avenir. Ils rencontrent des facteurs de complication qui sont, bien sûr, la prolifération des contenus en ligne et à la demande qui implique qu'il n'y ait plus besoin d'« émissions programmées », et la difficulté à trouver le moyen d'offrir aux consommateurs un système de navigation plus précis et personnalisé, couvrant toutes les formes de contenus disponibles, quelle qu'en soit l'origine.

De substantiels changements organisationnels sont nécessaires pour améliorer l'agencement des métadonnées destinées aux consommateurs. Mais il est souvent difficile de faire évoluer les habitudes de travail au sein du secteur professionnel de la radiodiffusion, qui déploie déjà de gros efforts pour migrer vers des systèmes numériques à base de fichiers. En outre, par suite de la mise en œuvre très rapide et exhaustive de normes internationales de métadonnées (telles que TV Anytime), un très grand nombre de « tags » peuvent être utilisés pour annoter un contenu et faciliter la recherche et la navigation.

2. Le problème des normes de métadonnées

Avec l'existence de telles normes, on pouvait s'attendre à ce que les innombrables producteurs de contenus, chaînes télévisées, radiodiffuseurs et autres acteurs appliquent des « tags » aux informations d'une façon homogène, permettant ainsi un échange d'informations fluide entre une vaste palette de services et de terminaux. Dans ce contexte, il est intéressant de remarquer que c'est sur Internet, du côté des consommateurs, que l'application de tags se produit le plus fréquemment et avec le plus de cohérence ; rappelons que ce sont les consommateurs qui ont été, par exemple, à l'origine de l'Internet Movie DataBase¹, et de la Compact Disc DataBase² (système permettant l'exploitation des métadonnées d'iTunes). Les contenus ainsi « étiquetés » par les utilisateurs (souvent appelés « folksonomies »)

1) <http://www.imdb.com>

2) <http://www.cddb.org>

offrent une perspective intéressante à étudier dans le contexte du travail énorme que constitue l'étiquetage des contenus. Les systèmes de ce type sont souvent auto-correcteurs, c'est-à-dire que les systèmes de recherche intelligents peuvent repérer des groupements de mots similaires pour interpréter l'intention de la recherche effectuée. Les métadonnées demeureront un élément clé dans l'obtention de résultats par les moteurs de recherche. Les innovateurs en matière de recherche sur les contenus vidéo s'efforcent d'aborder ces questions petit à petit, de façons multiples et variées. Mais en réalité, ils tentent de combler au moyen d'une technologie très sophistiquée ce qui manque encore sous une forme simple de métadonnées homogènes. Dans le prolongement du Web 2.0, centré sur les réseaux sociaux, du Web sémantique et de son développement à venir, il est fort probable que ce seront les consommateurs qui contribueront à la création unique de métadonnées qui devront ensuite être maintenues à jour d'une façon cohérente, peut-être à la façon de Wikipedia.

Aujourd'hui, le défi réside dans le fait que les grands propriétaires de contenus rejettent le concept (ou, peut-être, ne le comprennent pas) selon lequel les données décrivant leurs contenus devraient être disponibles gratuitement aux consommateurs et interopérables avec d'autres systèmes. Il s'agit essentiellement d'une question d'éducation : les métadonnées sont comparables à l'emballage d'un produit. Mais le « packaging des produits » sur le marché numérique relève davantage de la diffusion et du traitement des données (c'est une question concernant les technologies de l'information) que du « marketing » ou de la « publicité » tape-à-l'œil traditionnels auxquels on l'a toujours associé.

Il existe des normes internationales permettant une identification unique des œuvres audiovisuelles (par exemple l'ISAN, International Standard Audiovisual Number, www.isan.org) – et il est probablement de l'intérêt de tous que les œuvres soient enregistrées correctement, une fois pour toutes. L'enregistrement d'une œuvre audiovisuelle par l'ISAN est toutefois payant, tandis que le secteur musical a adopté une toute autre approche ; à l'instar du secteur du livre et des numéros ISBN (un système plus ancien, également payant), on n'imagine pas les grands labels ne pas attribuer un identifiant unique aux enregistrements. Si les conditions nécessaires étaient réunies, l'application de l'ISAN (ou de tout autre système d'identification unique) pourrait considérablement simplifier le problème. Mais même avec un système équivalant à un « code barre » ou un code universel des produits (système que nous connaissons tous), il resterait un défi : modifier les habitudes de travail de façon à intégrer ces codes uniques aux innombrables bases de données existantes !

Il reste aujourd'hui de nombreux silos d'information distincts qui reproduisent à l'envi les mêmes informations descriptives au sujet des contenus audiovisuels. Il peut s'agir d'agrégateurs de données, de propriétaires de contenus, de sociétés de droits d'auteur, de sociétés spécialisées dans l'étude des audiences, de guides de programmes télévisés et de bien d'autres acteurs. Si les synopsis des émissions peuvent varier en fonction de la langue utilisée, de l'auteur ou du promoteur de tel ou tel programme (comme le résumé descriptif d'un épisode de « Grey's Anatomy »), les autres détails « statiques » n'ont pas besoin d'être conservés dans des bases de données distinctes. Il s'agit en effet de métadonnées immuables : le titre original et le titre de l'épisode, les noms des acteurs, du réalisateur et du producteur, ainsi que l'année de production. Un identifiant unique a donc une utilité économique, car les efforts déployés actuellement pour tenir à jour ces bases de données finissent par faire doublon.

Les défis sont donc nombreux, et de nouvelles questions juridiques viennent désormais les compliquer. Avec la mise en œuvre à venir de la Directive services de médias audiovisuels de l'Union européenne (d'ici à décembre 2009), et face aux nombreuses autres initiatives européennes en matière de réglementation et de politique, des exigences nouvelles voient le jour concernant la labellisation obligatoire des contenus dans les services de télévision et de vidéo à la demande. Au Royaume-Uni, les nouvelles initiatives en faveur de la protection des mineurs, telles que le rapport Byron, vont aboutir à la création d'un Conseil britannique sur la sécurité des enfants sur Internet (en septembre 2008). Ses recommandations, qui ont reçu le plein soutien du Gouvernement britannique, illustrent déjà explicitement l'impact qu'auront les nouvelles exigences de labellisation des contenus sur les entreprises. Dans ce cas précis, les jeux vidéo doivent maintenant porter au Royaume-Uni le système de labellisation du British Board of Film Classification (BBFC – Conseil britannique de classification des films). Le BBFC fournit également une autre « métadonnée » intéressante qui prend la forme d'un synopsis et d'un avertissement sur le contenu, et décrit de façon relativement détaillée le contenu du jeu vidéo (ou du film, etc.). Si les autres initiatives d'autorégulation entreprises par les producteurs

de jeux vidéo, comme le système PEGI³, étaient jusqu'alors suffisantes et formaient le système de labellisation à employer au Royaume-Uni, l'introduction du système du BBFC dans le droit britannique n'est qu'un exemple de la façon dont les systèmes contradictoires de labellisation finiront par gêner la mise en œuvre de systèmes effectifs et cohérents de navigation dans les contenus. Leur introduction sera plus complexe encore pour les sociétés de diffusion de contenus qui opèrent sur différents territoires et devront se conformer à plusieurs régimes juridiques.

3. L'importance de la saisie cohérente des données

Les consommateurs ont aujourd'hui de plus en plus de mal à identifier les contenus présentant un réel intérêt parmi une série de plateformes. Les sociétés de médias et les entreprises d'électronique grand public tentent désormais de combler les manques au moyen de « guides spécialisés » et de services de « navigation dans les contenus ». Mais à ce jour, ces derniers ne fonctionnent correctement qu'au sein de ce que l'on appelle des « jardins murés », où les données relatives au EPG sont clairement délimitées (par exemple : 500 chaînes sur BSkyB, auxquelles s'ajoutent la VoD et Sky Anytime – cela constitue un ensemble de données gérable).

Toutefois, fusionner de tels « jardins murés » avec des contenus Internet est un exercice complètement différent. Sur Internet, les moteurs de recherche portant sur la vidéo emploient de nombreuses technologies pour identifier les contenus – ils identifient par exemple ce qui n'est **pas** présent au moyen de métadonnées homogènes (en analysant le son, les pistes de sous-titres, la reconnaissance faciale, etc.). Et là où des technologies de filigrane/d'empreinte digitale sont utilisées, ces dernières ne sont pas encore systématiquement reliées à un service de métadonnées au contenu fiable (c'est-à-dire renvoyant à un identifiant unique, quel qu'il soit) – en d'autres termes, elles constituent des silos d'information supplémentaires.

Les choses seraient bien plus simples si chaque contenu était identifié – par un seul identifiant – et pouvait ensuite être accessible à tous par voie numérique. Les contenus audiovisuels pourraient être décrits une seule fois, d'une façon comparable à l'identification et à la description unique de chaque mot dans un dictionnaire. Là aussi, certains éléments statiques ne changent pas (par exemple, la forme grammaticale : nom, verbe, etc.), à l'instar de ceux des contenus audiovisuels : titre, nom du réalisateur, pays d'origine, année de production, etc. ne sont pas non plus amenés à être modifiés. Le texte descriptif (subjectif) pourrait être traité au moyen de liens vers une archive principale (avec différentes versions linguistiques, différentes longueurs de synopsis, différentes perspectives rédactionnelles, etc.).

Le Web sémantique promet une nouvelle génération de services d'information. Les formats de balisage XML⁴ tels que TV Anytime⁵ requièrent de gros efforts pour être mis en place. Et ils ne constituent qu'une sorte de métadonnées parmi beaucoup d'autres. Pour obtenir un fonctionnement général harmonieux, il faudrait que tout le monde adopte cette norme et que la saisie des données soit homogène (imaginons que quelqu'un, en Allemagne, étiquette une émission en utilisant le terme « Komödie »⁶, tandis que le terme « sitcom » sera utilisé aux Etats-Unis pour décrire le même programme).

Pour chaque bon argument en faveur d'une telle évolution des métadonnées, de l'étiquetage et de la navigation dans les contenus, il y a aussi des contre-arguments. Les motivations commerciales qui pourraient potentiellement amener les propriétaires de contenus à ouvrir soudain leurs bases de données librement au monde entier passeraient par :

3) Pan European Game Information (www.pegi.info).

4) Le langage XML (Extensible Markup Language) est une manière flexible de créer des formats communs d'information et de partager tant le format que les données sur le Web, les Intranets et ailleurs.

5) www.tv-anytime.org

6) En français, « comédie ».

- une meilleure compréhension du rôle de « packaging des produits » que jouent les métadonnées dans un environnement IT (par opposition à leur fonction de marketing traditionnelle) ;
- une volonté et une capacité de trouver une solution interopérable pour la protection des contenus (par exemple, une gestion interopérable des droits numériques, ou DRM).

Pour l'heure, l'acquisition et l'utilisation de métadonnées professionnelles, présentables aux consommateurs et permettant une recherche de fond, demeurent des services payants, qu'il faut commander, sauf si l'on veut bien se contenter d'avoir recours aux moteurs de recherche.

4. Les futurs systèmes de navigation dans les contenus

En matière de recherche, Internet conserve l'avantage. Ses moteurs de recherche sont bien plus sophistiqués et plus à même de repérer rapidement (quoique de façon un peu brouillonne) un contenu grâce à la simple saisie d'un mot ou d'une expression. Plus significatif encore est le fait que la jeune génération est passée maître dans l'art de manipuler logiciels et services au quotidien pour trouver ce qu'elle veut sur Internet, le tout dans un environnement multitâches. Pour les opérateurs traditionnels de télévision à péage, le défi est évident ; et lorsque l'on compare les interfaces existant dans l'univers de la télévision payante à boîtier décodeur, l'écart est stupéfiant⁷.

Les produits de type « guides spécialisés » ont fait leur apparition, mais l'accès à des métadonnées cohérentes, à la fois dans les sources Internet et les sources radiodiffusées, demeure un défi considérable. L'avenir de la navigation dans les contenus dépend fortement de l'accès à une source abondante de métadonnées précises et homogènes telles que :

- une entrée originale,
- identifiée de façon unique,
- accessible à tous (professionnels et grand public),
- dont l'orthographe, la grammaire et la ponctuation sont corrects, multilingue,
- à laquelle sont attribués de nombreux tags,
- extensible – afin que les commentaires et descriptions (ainsi que les photos, les vidéos, etc.) générés par les utilisateurs puissent y être adjoints en utilisant l'identifiant unique comme point d'ancrage.

Il reste à voir comment le marché des métadonnées évoluera pour permettre à de nouveaux services de navigation dans les contenus de voir le jour. Les produits de type « guides spécialisés » existent, mais n'ont pas encore été mis en œuvre, car la question de l'accès à des métadonnées fiables, tant dans les sources des radiodiffuseurs que sur Internet, reste très délicate.

Même les sociétés d'agrégation de métadonnées les plus professionnelles sont incapables de suivre les exigences croissantes posées aujourd'hui par l'industrie (pochettes de DVD, photos, extraits vidéo, mots clés, systèmes internationaux de classification des films, avertissements quant au contenu, informations sur l'accessibilité aux personnes handicapées, etc.). Et les exigences en matière de labellisation vont s'accroître à mesure que les politiques de réglementation vont évoluer. L'avenir réside dans l'éducation, les incitations, les outils et le soutien du secteur public, afin de construire un environnement où l'économie numérique pourra s'épanouir. Dans un scénario de ce type, une gamme exceptionnelle de nouveaux produits et services trouvera alors sa place sur le marché.

7) Pour des exemples, consulter www.epgmetadata.com

Protection de la vie privée et quête du moteur de recherche audiovisuel parfait

Michael Zimmer
School of Information Studies
Center for Information Policy Research
Université de Wisconsin-Milwaukee

Introduction

Les moteurs de recherche sur Internet se sont imposés comme des outils omniprésents et essentiels permettant de naviguer dans la sphère toujours plus vaste de l'information en ligne. Le plus populaire d'entre eux, Google, résume dans le descriptif de sa mission l'objectif commun à tous les moteurs de recherche généralistes : « organiser l'information mondiale dans le but de la rendre accessible et utile à tous¹ ». A cet effet, Google, comme la plupart des fournisseurs d'outils du même type, s'efforce de créer « un moteur de recherche parfait apte à traiter et à comprendre la totalité des informations dans le monde² », qui saisit « exactement l'objet de chaque recherche pour fournir précisément les informations demandées³ ». La volonté de mettre au point un tel moteur de recherche a conduit à l'apparition d'outils puissants qui permettent d'accéder à des dizaines de milliards de sources d'information en ligne, traitent quotidiennement des millions de requêtes, et fournissent des résultats pertinents et souvent personnalisés – le tout, gratuitement.

Avec l'avènement des largeurs de bande à grande vitesse et des technologies numériques peu onéreuses, l'omniprésence croissante des webcams et des caméras incorporées aux téléphones portables, et l'existence de sites Internet innovants, destinés au stockage de contenus et à l'interaction sociale (tels que Flickr ou Facebook), le Web est de plus en plus une plateforme de création, de stockage et de partage de contenus multimédias générés par les utilisateurs. Par suite, les fournisseurs de moteurs de recherche s'efforcent d'accroître la puissance et la portée de leurs robots indexeurs, afin d'intégrer des images ainsi que des documents audio et vidéo dans leurs index interrogeables, mettent leurs propres plateformes à disposition pour le stockage, la localisation et le partage d'images et de vidéos, voire, se spécialisent, pour proposer exclusivement des contenus multimédias (à l'instar du projet de moteur de recherche européen Quaero). En résumé, le moteur de recherche parfait est en train de devenir au premier chef audiovisuel.

La puissance et l'omniprésence croissantes des moteurs de recherche ont suscité la controverse, et ont notamment fait naître des inquiétudes quant aux partis-pris des moteurs de recherche, à la censure (ou à son contournement), au non-respect du droit d'auteur, aux contrefaçons de marques commerciales, ainsi qu'à la liberté d'expression sur Internet. Mais la crainte la plus notable est peut-être celle

1) Google. (2005). Présentation de la société. Consulté le 3 mai 2006 sur <http://www.google.com/corporate/index.html>

2) *Google Press Center*. (7 juin 1999). *Google Receives \$25 million in equity funding* (Google reçoit 25 millions de dollars sous forme de fonds d'action), communiqué de presse, juin 1999. <http://www.google.com/press/pressrel/pressrelease1.html>, consulté le 18 août 2006.

3) Google. (2007). *Our philosophy*. Consulté le 27 mars 2007, sur : <http://www.google.com/intl/fr/corporate/tenthings.html>

qui concerne la protection de la vie privée face à ces moteurs de recherche puissants⁴. En s'efforçant de rendre les contenus audiovisuels aussi « accessibles et utiles à tous » que leurs cousins textuels, les puissants outils de recherche audiovisuels qui ont fait l'apparition ont ravivé les inquiétudes dans ce domaine, tant parmi les universitaires qu'au sein des organismes de réglementation⁵. En raison de la nature particulière des contenus audiovisuels, la quête d'un « moteur de recherche audiovisuel parfait » soulève des questions méritant une attention particulière en matière de protection la vie privée. Le présent article décrit certaines des dernières évolutions survenues dans les technologies de recherche audiovisuelle (telles que l'intégration d'images créées par des utilisateurs dans les index interrogeables, l'attribution d'abondants « tags » et de métadonnées, et les outils de reconnaissance faciale destinés aux utilisateurs) et met en lumière les menaces qu'elles inspirent quant à la protection de la vie privée.

1. La quête du moteur de recherche parfait

Depuis l'apparition des premiers moteurs de recherche, capables de servir d'interface entre les utilisateurs et les contenus en ligne, on a vu naître la quête d'un « moteur de recherche parfait », un outil capable d'indexer toutes les informations disponibles et de fournir rapidement des résultats pertinents⁶. Ce moteur de recherche parfait offrirait des résultats intuitifs, fondés sur les recherches antérieures effectuées par les utilisateurs ainsi que sur l'historique de leur navigateur⁷. Il saurait déterminer, par exemple, si une recherche sur les mots-clés « Paris Hilton » vise à localiser des chambres d'hôtel dans la capitale française ou à obtenir les derniers ragots concernant une jeune célébrité très en vue ; dans chacun de ces deux cas, il ne fournirait que les résultats adéquats. D'importantes motivations financières poussent bien sûr les sociétés conceptrices de moteurs de recherche à vouloir atteindre cette perfection de la recherche : le fait de proposer des résultats de recherche personnalisés pourrait fidéliser les utilisateurs à un service donné de moteur de recherche, ce qui accroîtrait leur exposition aux partenaires publicitaires de ce site et les inciterait à avoir recours à des services payants. De même, les moteurs de recherche peuvent augmenter leurs tarifs publicitaires dès lors qu'ils placent des publicités ciblées sous les yeux d'utilisateurs ayant des besoins et des centres d'intérêts liés aux produits vantés (par exemple, un internaute recherchant des informations en vue de voyager, et non des ragots sur une célébrité)⁸.

Le journaliste John Battelle⁹ décrit de la façon suivante le fonctionnement que pourrait adopter un moteur de recherche parfait :

Imaginez de pouvoir poser la question de votre choix et d'obtenir non seulement une réponse exacte, mais la réponse parfaite pour vous – une réponse qui correspond au contexte et à l'inten-

4) Voir par exemple Sullivan, D. (2 avril 2003). *Search privacy at Google & other search engines*. Consulté le 31 mars 2007 sur : <http://searchenginewatch.com/showPage.html?page=2189531> ; Associated Press. (17 juillet 2005). *Google Growth yields privacy fear*. Consulté le 31 mars 2007 sur : <http://www.wired.com/politics/security/news/2005/07/68235> ; Mills, E. (3 août 2005). *Google balances privacy, reach*. Consulté le 7 janvier 2007 sur :

http://news.com.com/Google+balances+privacy,+reach/2100-1032_3-5787483.html ; Zimmer, M. (2008). *Privacy on planet Google: Using the theory of "contextual integrity" to clarify the privacy threats of Google's quest for the perfect search engine*. *Journal of Business & Technology Law*, 3(1), 109-126 ; Zimmer, M. (2008) *The gaze of the perfect search engine: Google as an infrastructure of dataveillance*. Dans A. Spink, & M. Zimmer (Eds.), *Web searching: Multidisciplinary perspectives*, p. 77-99. Dordrecht, Pays-Bas : Springer.

5) Ainsi, en septembre 2007, la Jan van Eyck Academie (Maastricht, Pays-Bas) a organisé une table-ronde décisive sur le projet de recherche audiovisuelle européenne Quaero (voir : http://www.janvaneyck.nl/0_2_3_events_info/arc_07_quaero_thinktank.html). Elle a été suivie en avril 2008 par un atelier d'experts sur les défis en matière de réglementation que posent les moteurs de recherche audiovisuels, qui s'est tenu à l'Institut du droit de l'information, à l'université d'Amsterdam.

6) Voir Kushmerick, N. (23 février 1998). *The search engineers*. *The Irish Times*, p. 10 ; Andrews, P. (7 février 1999). *The search for the perfect search engine*. *The Seattle Times*, p. E1 ; Gussow, D. (4 octobre 1999). *In search of...* *St. Petersburg Times*, p. 13 ; Mostafa, J. (24 janvier 2005). *Seeking better Web searches*. Consulté le 30 janvier 2005 sur : http://www.sciam.com/print_version.cfm?articleID=0006304A-37F4-11E8-B7F483414B7F0000

7) Pitkow, J., Schütze, H., Cass, T., Turnbull, D., Edmonds, A., & Adar, E. (2002). *Personalized search*. Communications de l'ACM, 45(9), 50-55 ; Teevan, J., Dumais, S. T., & Horvitz, E. (2005). *Personalizing search via automated analysis of interests and activities*. Actes de la 28^e conférence annuelle internationale de l'ACM SIGIR consacrée à recherche et au développement dans la localisation de l'information, 449-456.

8) Hansell, S. (26 septembre 2005). *Microsoft plans to sell search ads of its own*. *The New York Times*, p. C1, C8.

9) Battelle, J. (8 septembre 2004). *Perfect search*. *Searchblog*. Consulté le 16 mai 2006 sur : <http://battellemedia.com/archives/000878.php>

tion de votre question, une réponse modulée en fonction de votre personne et de la raison pour laquelle vous posez cette question. Le moteur de recherche offrant cette réponse est capable de compiler toutes les connaissances disponibles sur le sujet dans le monde entier, qu'elles soient sous forme de texte, de vidéos ou de fichiers audio. Il est en mesure de distinguer une requête simple (« Qui était le troisième Président des Etats-Unis ? ») d'une demande plus subtile (« Dans quelles circonstances le troisième Président des Etats-Unis renia-t-il sa conception de l'esclavage ? »).

Ce moteur de recherche parfait dispose également d'une mémoire parfaite. Il sait quelles pages vous avez consultées et peut faire la différence entre une exploration (visant à découvrir quelque chose de nouveau) et la récupération de résultats antérieurs (visant à retrouver une page déjà consultée précédemment).

A la question : « A quoi ressemblerait un moteur de recherche parfait ? », Sergey Brin, de Google, a répondu : « Au cerveau de Dieu »¹⁰. Pour atteindre un tel idéal d'omniprésence et d'omniscience, le moteur de recherche parfait doit disposer à la fois de la « portée parfaite » (pour accéder à l'ensemble des informations disponibles sur Internet) et de la « mémoire parfaite » (pour fournir des résultats personnalisés et pertinents, fondés sur les habitudes déjà observées de chaque utilisateur).

1.1. La portée parfaite

Pour atteindre la portée nécessaire à la concrétisation du projet de Recherche 2.0 (« Search 2.0 »), les moteurs de recherche en ligne constituent de gigantesques index répertoriant le contenu de la Toile. Au-delà des seules pages Internet de type HTML, les fournisseurs de moteurs de recherche ont répertorié une multitude de médias disponibles sur Internet, notamment des images, des fichiers vidéo et PDF, et d'autres documents informatiques. Ainsi, en 2005, Yahoo! affirmait avoir indexé plus de 20 milliards d'éléments, dont plus de 19,2 milliards de pages Web, 1,6 milliard d'images et plus de 50 millions de fichiers audio et vidéo¹¹. De plus, les moteurs de recherches tendent de plus en plus à numériser le « monde matériel », ajoutant à leurs index grandissants le contenu de livres à succès, de bibliothèques universitaires, de cartes et d'images satellites. Outre leurs index interrogeables, d'une grande variété et en constante expansion, les moteurs de recherche actuels visent aussi la « portée parfaite » en élaborant divers outils et services censés aider les utilisateurs à organiser et à utiliser des informations dans des contextes qui ne relèvent pas de la recherche « traditionnelle » sur Internet. Il peut s'agir de plateformes de communication, de réseaux sociaux, d'outils de gestion des données personnelles ou financières, de sites d'achats et de recherche de produits, de services de gestion de fichiers informatiques et de navigation améliorée sur Internet.

En réunissant ces deux aspects de la « portée parfaite » – des index interrogeables toujours plus vastes et des produits variés pour l'organisation des informations – le moteur de recherche parfait permet aux utilisateurs de chercher, de trouver et d'interagir avec la quasi-totalité des formes d'information dont ils ont besoin au quotidien. La portée des moteurs de recherche parfaits permet donc aux utilisateurs de chercher et d'accéder à la quasi-totalité des contenus disponibles sur Internet, mais aussi de communiquer, de visiter des sites d'achat et d'organiser leur vie, sur le Web et hors ligne.

1.2. La mémoire parfaite

Pour compléter cette *portée* parfaite, les fournisseurs de moteurs de recherche ambitionnent de conserver une *mémoire* parfaite de chaque utilisateur, afin de personnaliser les services et les publicités qui lui sont destinés. Pour obtenir cette mémoire parfaite, les moteurs de recherche en ligne doivent être en mesure d'identifier et de comprendre les attentes intellectuelles, les besoins et les désirs des utilisateurs lorsqu'ils recherchent des informations sur Internet. Pour déterminer le contexte et la finalité d'une recherche sur les termes « Paris Hilton », par exemple, le moteur de recherche parfait doit savoir si l'utilisateur a récemment manifesté le désir de se rendre en France ou, au contraire, s'il consacre un temps important à la recherche de potins sur les célébrités sur Internet.

10) Ferguson, C. (2005). *What's next for Google? Technology Review*, 108(1), 38-46.

11) Mayer, T. (8 août 2005). *Our blog is growing up – and so has our index. Yahoo! Search Blog*. Consulté le 25 novembre 2006 sur : <http://www.ysearchblog.com/archives/000172.html>

Le moyen privilégié de personnaliser les résultats des recherches est de s'appuyer sur les habitudes et l'historique de recherche de l'utilisateur¹². Pour recueillir l'historique des recherches des utilisateurs, la plupart des moteurs de recherche utilisent les logs de leurs serveurs, qui enregistrent chaque demande de recherche sur Internet traitée par ces serveurs, l'adresse IP soumettant la requête, les pages visualisées et les résultats consultés¹³. Les moteurs de recherche ont aussi largement recours aux cookies pour différencier les utilisateurs et suivre leur activité d'une session à l'autre ; en outre, ils incitent de plus en plus à la création de comptes d'utilisateurs, ce qui leur permet de faire le lien entre tel utilisateur et son activité en ligne. Cette collecte d'informations concernant les utilisateurs est justifiée sous l'angle d'une amélioration des recherches effectuées. Google affirme par exemple : « Ces informations sont utilisées pour améliorer la qualité de nos services et à des fins commerciales¹⁴ », tandis que le moteur de recherche Ask.com avance également les raisons économiques le poussant à constituer cette mémoire parfaite, dans le cadre de la poursuite de l'objectif d'une recherche parfaite : « Nous recueillons... des informations de façon anonyme, afin d'améliorer la qualité globale du temps passé en ligne, ce qui inclut le suivi et l'amélioration de nos produits, la publicité ciblée et la monétisation des mots-clés à potentiel commercial¹⁵ ». Au vu de la croissance de la gamme de produits et de services proposée par les fournisseurs de moteurs de recherche, les logs de leurs serveurs sont susceptibles de contenir bien plus que les seules requêtes de recherches saisies par les internautes¹⁶. Les recherches portant sur différentes sortes de contenus en ligne – images, actualité, vidéos, livres, recherches universitaires, produits, musique, etc. – sont très certainement enregistrées, tout comme les liens consultés et les statistiques d'utilisation provenant de services qui ne relèvent pas de la recherche, mais qui sont proposés par les mêmes prestataires (messagerie électronique, groupes de discussion, services de cartographies, etc.).

En résumé, la quête du moteur de recherche parfait a conduit les fournisseurs d'outils de recherche à étendre la portée de leurs services pour indexer et rendre accessible des sources d'information très variées (tant en ligne qu'hors ligne), mais aussi à améliorer la mémorisation des habitudes, des besoins et des souhaits de chaque utilisateur. Si l'on tient compte des mesures prises pour obtenir la portée et la mémoire parfaites évoquées plus haut, les moteurs de recherche semblent se rapprocher de plus en plus de cet idéal d'omniscience et d'omniprésence, et être assurés de « devenir la force parfaite du 21^e siècle, omnipotente, capable de tout voir et de tout savoir¹⁷ ».

2. Moteur de recherche parfait et confidentialité

L'ambition de mettre au point le moteur de recherche parfait – promettant exhaustivité, variété et pertinence – n'est pas sans danger. Le chemin conduisant à sa concrétisation est jalonné d'inquiétudes, concernant notamment les partis-pris, la censure (ou son contournement), le non-respect du droit d'auteur, les contrefaçons de marques commerciales, ainsi que la liberté d'expression¹⁸. Mais les inquiétudes les plus prégnantes sont celles qui concernent la préservation de la confidentialité dans ce contexte¹⁹. En effet, tant la portée que la mémoire nécessaires pour obtenir cet outil parfait ont des conséquences particulières dans le domaine de la protection des données privées.

12) Voir par exemple Speretta, M. (2000). *Personalizing search based on user search histories*. University of Kansas ; Pitkow, J., Schütze, H., Cass, T., Turnbull, D., Edmonds, A., & Adar, E. (2002). *op.cit.* ; Teevan, J., Dumais, S. T., & Horvitz, E. (2005). *op.cit.*

13) Voir par exemple Google. (2005). *Google Privacy faq*. Consulté le 3 mai 2006 sur : http://www.google.com/privacy_faq.html ; IAC Search & Media. (13 juillet 2005). *Privacy policy for Ask.com*. Consulté le 6 janvier 2007 sur : <http://sp.ask.com/en/docs/about/privacy.shtml> ; Yahoo! (11 novembre 2006). *Yahoo! Privacy policy*. Consulté le 6 janvier 2007 sur : <http://info.yahoo.com/privacy/us/yahoo/details.html>

14) Google. (2005) *Google Privacy faq*. *Op.cit.*

15) IAC Search & Media. *Op.cit.*

16) Wikipedia énumère plus de cent produits et services différents proposés par Google, et plus de cinquante pour Yahoo !

17) Ayers, C. (1^{er} novembre 2003). *Google: Could this be the new God in the machine?* *The Times*, p. 4.

18) Voir par exemple Grimmelmann, J. (à paraître). *The structure of search engine law*. *Iowa Law Review* ; Introna, L. & Nissenbaum, H. (2000). *Shaping the Web: Why the politics of search engines matters*. *The Information Society*, 16(3), 169-185 ; Elkin-Koren, N. (2001). *Let the crawlers crawl: On virtual gatekeepers and the right to exclude indexing*. *University of Dayton Law Review*, 26, 180-209 ; Gasser, U. (2006). *Regulating search engines: Taking stock and looking ahead*. *Yale Journal of Law & Technology*, 9, 124-157 ; Goldman, E. (2006). *Search engine bias and the demise of search engine utopianism*. *Yale Journal of Law & Technology*, 188-200.

19) Voir par exemple Sullivan, D. *op.cit.* ; Associated Press. *op.cit.* ; Mills, E. *op.cit.* ; Zimmer, M. (2008) *Privacy on planet Google*. *Op.cit.* ; Zimmer, M. (2008) *The gaze of the perfect search engine*. *Op.cit.*

2.1. Données personnelles et portée parfaite

Pour offrir une « recherche parfaite », les moteurs de recherche doivent indexer autant de pages Internet et de ressources en ligne qu'ils le peuvent, afin de constituer la plus grande base de données possible de résultats potentiels de recherche. Parmi les milliards de pages indexées par les moteurs de recherche, on trouve notamment des sites Internet contenant des informations personnelles sur des individus – tels que des pages personnelles (actives ou abandonnées), des posts sur des forums de discussion, des CV en ligne, des procès-verbaux de réunions publiques, des dossiers d'impôts fonciers ainsi que des documents judiciaires. Rares sont les gens qui ne sont pas touchés par le « long bras du robot indexeur de Google », explique le journaliste Neil Swidey en évoquant le champ très étendu de l'index de Google :

C'était peut-être une farce stupide dans une association d'étudiants, ou un post imprudent publié dans un newsgroup en ligne quand vous étiez à la fac. Peut-être vous êtes-vous un jour lancé dans une tirade lors d'un conseil municipal, ou avez-vous signé une pétition sans prendre le temps de la lire. Ou peut-être avez-vous subi un divorce difficile. Vous pensiez que ces chapitres de votre vie étaient définitivement terminés. Mais pas pour Google.

Si ce « bagage » embarrassant était déjà de fait public pour une grande partie, il n'était en réalité accessible qu'à des personnes faisant preuve d'une intrépidité professionnelle ou d'une curiosité extrême. Désormais, dans les états qui ont mis en ligne leurs archives judiciaires, vous pourrez lire tous les détails sordides du divorce de votre voisin d'un simple clic sur Google, sans plus d'efforts que pour consulter vos courriels²⁰.

Une « *vanity search* » (opération qui consiste à chercher son propre nom sur Internet) peut révéler une quantité surprenante d'informations privées : avis d'imposition, documents judiciaires, acte de mariage, acte de vente et informations concernant l'inscription sur les listes électorales, par exemple. Il est devenu courant de « googler » l'inconnu(e) avec lequel (laquelle) on a un rendez-vous galant²¹. Près d'un internaute sur quatre a déjà cherché sur Internet des informations sur ses collègues de travail ou ses relations d'affaires²², et les employeurs « googlent » leurs employés potentiels avant de prendre des décisions de recrutement²³. En moins d'une heure, un journaliste a découvert une série de détails privés concernant la vie d'Eric Schmidt, directeur général de Google :

- Schmidt révèle peu de choses sur lui-même sur son site personnel. Mais trente minutes sur le moteur de recherche Google suffisent pour apprendre que la fortune de Schmidt, âgé de 50 ans, était estimée à USD 1,5 milliard environ l'an dernier. Au début de l'année, il a gagné près de USD 90 millions en vendant des actions Google, puis au moins USD 50 millions en cédant des titres au cours des deux derniers mois, l'action Google ayant bondi à plus de USD 300.
- Il vit avec son épouse Wendy dans la ville cossue d'Atherton, en Californie où, il y a cinq ans, lors d'une levée de fonds politique à USD 10 000 la place, le candidat à la présidence Al Gore et son épouse Tipper ont dansé ensemble tandis qu'Elton John chantait à pleins poumons « Bennie and the Jets ».
- Schmidt a également arpenté le désert du Nevada dans le cadre du festival artistique Burning Man ; c'est un pilote amateur passionné.^{24/25}

20) Swidey, N. (2 février 2003). *A nation of voyeurs: How the Internet search engine Google is changing what we can find out about one another - and raising questions about whether we should.* *The Boston Globe Sunday Magazine*, p. 10.

21) Lobron, A. (5 février 2006). *Googling your Friday-night date may or may not be snooping, but it won't let you peek inside any souls.* *The Boston Globe Magazine*, p. 42.

22) Sharma, D. (21 octobre 2004). *Is your boss Googling you?* Consulté le 6 janvier 2007 sur : http://news.com.com/Is+your+boss+Googling+you/2100-1038_3-5421210.html

23) Weiss, P. (19 mars 2006). *What a tangled Web we weave : Being googled can jeopardize your job search.* *New York Daily News*. Consulté le 7 janvier 2007.

24) Mills, E. (3 août 2005). *Op.cit.*

25) De façon ironique, Google a puni CNET pour avoir publié des informations privées au sujet de Schmidt – trouvées via son propre moteur de recherche – en refusant pendant un an de répondre à quelque journaliste que ce soit de ce service de presse. Face à la critique du grand public, Google a mis fin à ce boycott au bout de deux mois.

L'accroissement de l'exhaustivité et de la variété des index interrogeables dans le but d'atteindre une « portée parfaite » a également amoindri le sentiment que pouvaient avoir les utilisateurs que leur vie privée était protégée par « l'ombre »²⁶ dans laquelle elle se trouvait. Avec les moteurs de recherche, il est désormais facile de trouver et d'accumuler un nombre considérable d'informations sur des particuliers – informations qui étaient autrefois difficiles à localiser, et plus encore à recouper. Résultat, il est de plus en plus aisé d'exploiter une quantité croissante d'informations à caractère privé disponibles en ligne, que ce soit pour se livrer à du harcèlement électronique²⁷, pour agréger ou explorer des données privées réparties dans des bases de données hétérogènes²⁸, ou pour établir des dossiers numériques au sujet des individus²⁹. Herman Tavani, spécialiste de l'éthique, a justement évoqué la facilité avec laquelle des informations personnelles peuvent être rassemblées systématiquement, puis agrégées et analysées par des moteurs de recherche sur Internet :

Du fait que les différents newsgroups contiennent des liens vers des informations postées par des particuliers, ils peuvent fournir aux utilisateurs de moteurs de recherche un aperçu très complet des centres d'intérêt et des activités de ces personnes³⁰. Il s'ensuit que toutes les informations à caractère personnel apparaissant actuellement sur les sites Internet accessibles aux moteurs de recherche n'y ont pas nécessairement été publiées par les personnes concernées, ou que leur publication n'a pas été explicitement autorisée par ces personnes³¹.

Un individu peut ignorer que son nom est inclus dans une ou plusieurs base(s) de données accessible(s) aux moteurs de recherche – sans parler du fait que tout le monde n'a pas une connaissance approfondie quant au fonctionnement des moteurs de recherche et leur capacité à localiser des informations personnelles issues de sources très variées.

C'est John Battelle qui résume le mieux l'anxiété née de la « portée parfaite » atteinte par les moteurs de recherche :

« Que faire, lorsque des informations dont nous savons qu'elles devraient être publiques deviennent... réellement publiques ? ... Que se passe-t-il, lorsque le moindre élément public à votre sujet (d'une mention dans le bulletin d'information de votre classe de CE1 – qui est désormais en ligne, bien sûr – aux délires vengeurs d'un amoureux éconduit) permet de suivre votre nom à la trace pour l'éternité³² ? »

2.2. Données personnelles et mémoire parfaite

Parallèlement à ces craintes, liées aux données privées et concernant la portée de plus en plus parfaite du moteur de recherche parfait, d'autres inquiétudes voient le jour quant à la capacité des moteurs de recherche à réunir et à agréger une vaste gamme d'informations d'ordre personnel et intellectuel au sujet de leurs utilisateurs, dans le but de constituer une « mémoire parfaite ». La plus évidente d'entre elles porte sur le suivi des requêtes soumises individuellement et la possibilité de les relier à des adresses IP particulières. Cette pratique, qui consiste à recueillir et à conserver les données relatives aux recherches effectuées dans le but d'obtenir une « mémoire parfaite », a suscité l'intérêt du grand public en janvier 2006 ; on apprenait alors que le ministère américain de la Justice avait demandé à un juge fédéral de contraindre le moteur de recherche Google à lui remettre les relevés des

26) Swidey, N. (2 février 2003). *op. cit.* ; Ramasastry, A. (12 mai 2005). *Can we stop zabasearch — and similar personal information search engines? When data democratization verges on privacy invasion*. Consulté le 12 juin 2006 sur : <http://writ.news.findlaw.com/ramasastry/20050512.html>

27) Tavani, H. & Grodzinsky, F. (2002). *Cyberstalking, personal privacy, and moral responsibility*. *Ethics and Information Technology*, 4(2), 123-132.

28) Garfinkel, S. (2000). *Database nation: The death of privacy in the 21st century* ; 1^{re} édition. Sebastopol, CA : O'Reilly.

29) Solove, D. (2004). *The digital person: Technology and privacy in the information age* ; Ex machina. New York : New York University Press.

30) Une recherche sur mon nom révèle des posts (que j'avoue avoir oubliés) sur des forums de discussion Usenet datant du début des années 1990, sur des sujets allant du droit à l'avortement au tableur Lotus 1-2-3, en passant par le catholicisme, le féminisme et le marketing.

31) Tavani, H. T. (2005). *Search engines, personal information and the problem of privacy in public*. *International Review of Information Ethics*, 3, 39-45.

32) Battelle, J. (2005). *The search: How Google and its rivals rewrote the rules of business and transformed our culture*. New York : Portfolio, p. 193.

recherches effectuées par plusieurs millions de ses utilisateurs³³. Google a refusé, mais trois de ses concurrents, America Online (AOL), Microsoft et Yahoo! se sont pliés à des assignations du même type délivrées par le gouvernement et portant sur leurs archives de recherches³⁴. Un peu plus tard, la même année, AOL a publié plus de 20 millions de requêtes soumises par 658 000 de ses utilisateurs, souhaitant alimenter ainsi la recherche universitaire sur l'analyse des requêtes soumises aux moteurs de recherche³⁵. Malgré les efforts d'AOL pour rendre ces données anonymes, les utilisateurs restaient identifiables à partir de leur seul historique de recherche, qui comprenait des mots-clés correspondant à leur nom, des numéros de sécurité sociale, des adresses, des numéros de téléphone et d'autres informations permettant de les identifier personnellement³⁶.

Ces affaires ont révélé au grand public les pratiques de rétention des requêtes de recherche et suscité l'anxiété de nombreux utilisateurs quant à l'existence d'un contrôle systématique de leurs activités de recherche d'informations sur Internet³⁷ ; de grands médias d'information ont été incités à enquêter et à publier des articles sur les informations recueillies quotidiennement par les moteurs de recherche au sujet de leurs utilisateurs³⁸. De leur côté, divers groupes de pression ont critiqué le fait que les moteurs de recherche en ligne soient capables de pister et de collecter dans de telles proportions des requêtes de recherche, souvent sans que les utilisateurs en aient conscience³⁹, tandis que les organismes de réglementation européens et américains se sont mis à enquêter sur ces pratiques et ces politiques de rétention d'informations⁴⁰.

L'importance accrue accordée par le grand public à l'archivage des requêtes de recherche des utilisateurs dans les logs des serveurs s'accompagne d'une nouvelle inquiétude, au sujet de la capacité des fournisseurs de moteurs de recherche de contrôler et d'agrèger les activités menées via l'ensemble de leurs produits et de leurs services⁴¹. En encourageant la création de comptes d'utilisateurs, la promotion croisée des services afin que les utilisateurs ne quittent pas le site Internet du moteur de recherche, et l'utilisation de cookies persistants au sein de tous les services, les moteurs de recherche s'assurent la possibilité de recueillir et d'agrèger tout un éventail d'informations à caractère personnel et intellectuel au sujet de leurs utilisateurs. Ces informations ne se limitent pas aux mots-clés sur lesquels portent les recherches, mais comprennent aussi les articles d'actualité qu'ils consultent, leurs centres d'intérêt, les blogs qu'ils fréquentent, les livres qu'ils apprécient, ainsi que d'autres détails très personnels quant à l'identité de ces personnes, leurs orientations politiques, leur état de santé, leur vie sexuelle, leur religion, l'état de leurs finances et leurs préférences d'achat⁴². Le résultat est ce que John Battelle appelle une « base de données d'intentions » :

« Ces informations représentent, sous une forme agrégée, un réceptacle des intentions du genre humain – une immense base de données de désirs, de besoins, de souhaits et de préfé-

- 33) Hafner, K., & Richtel, M. *Google resists U.S. subpoena of search data*. *The New York Times*, p. A1, C4 ; Mintz, H. (16 janvier 2006). *Feds after Google data: Records sought in U.S. quest to revive porn law*. *San Jose Mercury News*. Consulté le 19 janvier 2006 sur : <http://www.siliconvalley.com/mld/siliconvalley/13657386.htm>.
- 34) Hafner, K., & Richtel, M. (20 janvier 2006). *Google resists U.S. Subpoena of search data*. *The New York Times*, p. A1, C4.
- 35) Hansell, S. (8 August 2006). *AOL removes search data on vast group of Web users*. *The New York Times*, p. C4.
- 36) McCullagh, D. (7 août 2006). *Aol's disturbing glimpse into users' lives*. Consulté le 3 décembre 2006 sur : http://news.com.com/AOLs+disturbing+glimpse+into+users+lives/2100-1030_3-6103098.html?tag=st.num
- 37) Barbaro, M., & Zeller Jr, T. (9 août 2006). *A face is exposed for AOL searcher no. 4417749*. *The New York Times*, p. A1 ; Hansell, S. (8 août 2006). *op.cit.* ; McCullagh, D. (7 août 2006). *Aol's disturbing glimpse into users' lives*. Consulté le 3 décembre 2006 sur : http://news.com.com/AOLs+disturbing+glimpse+into+users+lives/2100-1030_3-6103098.html?tag=st.num
- 38) Glasner, J. (5 avril 2005). *What search sites know about you*. Consulté le 2 août 2006 sur : <http://www.wired.com/news/privacy/0,1848,67062,00.html> ; Ackerman, E. (19 août 2006) *What do Google, Yahoo, AOL and Microsoft's MSN know about you?* *San Jose Mercury News*.
- 39) Voir par exemple Electronic Frontier Foundation. (2007). *Privacy and search engines*. Consulté le 28 juillet 2007 sur <http://www.eff.org/Privacy/search/> ; Privacy International. (9 juin 2007). *A race to the bottom: Privacy ranking of Internet service companies*. Consulté le 10 juillet 2007 sur : <http://www.privacyinternational.org/article.shtml?cmd%5B347%5D=x-347-553961>
- 40) Associated Press. (25 mai 2007). *EU data privacy officers launch investigation into Google's Internet search engine*. Consulté le 28 juillet 2007 sur : <http://www.iht.com/articles/ap/2007/05/25/business/EU-FIN-EU-Google-Privacy-Probe.php> ; Lohr, S. (29 mai 2007). *Google Deal said to bring U.S. Scrutiny*. Consulté le 27 juillet 2007 sur : <http://www.nytimes.com/2007/05/29/technology/29antitrust.html>
- 41) Glasner, J. (5 avril 2005). *op. cit.* ; Ackerman, E. (19 août 2006). *op.cit.* ; Waters, D. (31 mai 2007). *What search engines know about us*. Consulté le 2 avril 2008 sur : <http://news.bbc.co.uk/2/hi/technology/6700997.stm>
- 42) Zimmer, M. (2008) *The gaze of the perfect search engine*. *Op.cit.*

rences, qui peuvent être découverts, cités, archivés, pistés et exploités à toute sorte de fins. Il n'y a jamais eu pareil monstre dans l'histoire de la culture, mais il est pratiquement garanti qu'il connaîtra dorénavant une croissance exponentielle. Cet artefact nous apprend des choses extraordinaires sur l'identité de notre culture et sur ce que nous voulons⁴³ ».

Si beaucoup de nos habitudes quotidiennes (l'utilisation de cartes de crédit, de DAB, de téléphones portables ou d'appareils de péage) laissent d'innombrables « empreintes virtuelles » de nos activités, la quête de la mémoire parfaite, nécessaire pour élaborer le moteur de recherche parfait, entraîne la mise en place d'une infrastructure massive de surveillance des données, qui dispose du pouvoir de contrôler et d'enregistrer les historiques de recherche des utilisateurs, les courriels, les posts des blogs et les habitudes de navigation générales ; tout ceci fournit « un excellent aperçu de ce que *pense* une personne, et pas seulement de ce qu'elle fait⁴⁴ ».

3. Quand la recherche parfaite devient audiovisuelle

Avec la prolifération des images numériques et des contenus vidéo, sur Internet et sur le Web, les fournisseurs de moteurs de recherche s'efforcent depuis longtemps d'inclure des contenus audiovisuels dans leurs index interrogeables. En 1995, AltaVista est le premier moteur de recherche à proposer la recherche d'images et de contenus audio et vidéo. Cette caractéristique est bientôt reprise par les autres grands moteurs de recherche de l'époque, notamment Lycos, HotBot et Infoseek. Ces premiers services de recherche audiovisuelle incluent l'accès à des fichiers multimédia, essentiellement sur la base des extensions des noms de fichiers (.gif, .avi, .au) et de références textuelles extraites de documents contextuels (nom du fichier, métadonnées, éléments de texte sous forme de lien hypertexte ou de mots entourant le fichier)⁴⁵. Aujourd'hui, si certains moteurs de recherche spécialisés subsistent (Picsearch, par exemple, permet d'effectuer une recherche parmi plus de deux milliards d'images), les services de recherche audiovisuelle sont dominés par les grands moteurs de recherche généralistes. Google, Yahoo, Microsoft et Ask offrent tous la possibilité de chercher des images et des fichiers vidéo, en utilisant les mêmes méthodes de localisation que celles mises au point par AltaVista il y a plus de dix ans.

Bien que la recherche audiovisuelle demeure encore fortement dépendante des métadonnées pour identifier le contenu des images et les localiser, des innovations ont vu le jour dans ce domaine – notamment des améliorations dans les services de recherche audiovisuelle, qu'elles soient de nature technologique ou destinées aux réseaux sociaux. Le développement des sites Web de partage de photos a entraîné une explosion de la quantité d'images personnelles disponibles en ligne, Flickr jouant ici un rôle prépondérant. Si ce site constitue une plateforme puissante, d'utilisation facile, pour mettre en ligne et partager des photos, la grande innovation de Flickr a été d'offrir aux utilisateurs la possibilité d'attribuer des « tags » à leurs photos, avec leurs propres étiquettes descriptives, ce qui permet de localiser des images liées à un thème donné (lieu ou sujet). Suite à son rachat par le moteur de recherche Yahoo, la base de données de Flickr, comptant plus de 100 millions de photos, a rapidement été intégrée aux résultats de recherche d'images de Yahoo, ce qui a fortement accru la portée de l'offre de recherche audiovisuelle de ce fournisseur⁴⁶.

A partir des tags attribués aux photos et créés par l'auteur de l'image au sein de Flickr, Yahoo a pu accroître la précision de son moteur de recherche et de localisation d'images : plutôt que d'utiliser le texte entourant la photo ou le simple nom de l'image pour déterminer si elle correspond à une requête particulière, Yahoo peut désormais tirer parti des nombreux tags placés délibérément par un utilisateur pour caractériser cette image. Google a rapidement mis au point sa propre méthode pour améliorer les résultats en matière de recherche d'images : le « Google Image Labeler »⁴⁷. Les utilisateurs sont groupés

43) Battelle, J. (13 novembre 2003). *The database of intentions*. Searchblog. Consulté le 16 mai 2006 sur : <http://battellemedia.com/archives/000063.php>

44) Hinman, L. (2005). *Esse est indicato in Google: Ethical and political issues in search engines*. *International Review of Information Ethics*, 3, 19-25, p. 23.

45) Schwartz, C. (1998). *Web search engines*, *Journal of the American Society for Information Science*, 49(11), 973-982.

46) Yahoo! (26 juin 2007). *Flickr-izing image search*. Searchblog Yahoo!. Consulté le 10 septembre 2007 sur : <http://www.ysearchblog.com/archives/000466.html>

47) Lenssen, P. (1^{er} septembre 2006). *Google image labeler*. Consulté le 11 juillet 2007 sur : <http://blogoscoped.com/archive/2006-09-01-n44.html>

deux à deux de façon aléatoire et se voient présenter une image issue de la base de données de Google. Chacun établit des étiquettes pour décrire l'image, sans voir les propositions de l'autre utilisateur. S'ils proposent le même mot dans les 120 secondes, ils gagnent des points. Si cette option est divertissante pour les utilisateurs, elle constitue surtout pour Google un moyen ingénieux de s'assurer que ses images sont associées aux mots-clés les plus adaptés.

Parallèlement à l'utilisation croissante de tags générés par des utilisateurs et des réseaux communautaires pour faciliter l'identification et la localisation d'images, des avancées concernant les technologies automatisées de localisation des images par le contenu ont nettement amélioré la recherche audiovisuelle⁴⁸. Au lieu de se reposer sur des métadonnées créées par des utilisateurs en chair et en os, les systèmes de localisation par le contenu tentent de reconnaître et de cataloguer différentes formes, couleurs et textures au sein des images, opérant par exemple une distinction automatique entre la photo d'un chien marron et celle d'une pomme rouge. La possibilité d'identifier automatiquement et de distinguer des visages au sein des systèmes de localisation d'images a récemment suscité un intérêt accru, tant à des fins policières que commerciales⁴⁹. Ces systèmes de reconnaissance faciale sont de plus en plus mis en œuvre dans la sphère du consommateur et intégrés aux moteurs de recherche sur Internet. Ainsi, après avoir acquis Neven Vision, une société qui développe des technologies visant à détecter et identifier des objets ainsi que des personnes dans des images⁵⁰, Google a amélioré son service *Image Search* pour permettre de restreindre les résultats aux seules images contenant des visages. Ainsi, une recherche simple sur le terme « Zimmer » dans *Image Search* permet d'obtenir des résultats variés, depuis une voiture exotique portant ce nom, à des photos d'intérieurs quelconques (« Zimmer » étant le terme allemand pour « pièce ») et des photos de particuliers. Mais en sélectionnant l'option « visages » dans les paramètres avancés du service *Image Search*, on peut se limiter aux résultats contenant des visages humains.

Divers services de recherche audiovisuelle ont récemment vu le jour, qui se concentrent exclusivement sur la reconnaissance faciale dans les images. Lancé début 2006, le service de recherche de photos Riya propose une technologie de reconnaissance faciale et textuelle pour identifier les visages au sein de la collection de photos numériques d'un utilisateur et leur attribuer des « tags » appropriés⁵¹. Une fois que l'utilisateur a rôdé le système manuellement en attribuant un tag au visage d'une même personne dans un certain nombre de photos, le logiciel prend le relais et peut désormais automatiquement scanner, identifier et attribuer un tag à cette personne dans toute future photo de son visage ajoutée à la photothèque de cet utilisateur⁵². Lors des deux premiers jours suivant son lancement, un million de photos avaient déjà été téléchargées vers le serveur de Riya ; puis sept millions de photos dans les sept premières semaines⁵³. Fort de ce succès, Riya a annoncé son projet de développer encore sa technologie de reconnaissance faciale au-delà des collections personnelles des utilisateurs, pour l'étendre à toute image sur le Web. Avec ce futur « moteur de recherche visuel », les utilisateurs pourront soumettre une image à Riya (un motif particulier de tapis, un sac à main vu sur une photo, ou même le visage d'une personne), et ce dernier fournira des images similaires issues de tout le Web⁵⁴.

La société suédoise Polar Rose prévoit des innovations du même type, en utilisant la reconnaissance faciale pour la localisation d'images⁵⁵. Plutôt que d'attendre que ses utilisateurs téléchargent leur photothèque personnelle vers son serveur pour analyse (selon le modèle de Riya), Polar Rose prévoit de fournir un module d'extension (plugin) pour les navigateurs, qui identifiera les visages sur les

48) Voir par exemple Lew, M., Sebe, N., Djeraba, C., & Jain, R. (2006). *Content-based multimedia information retrieval: State of the art and challenges*. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1), 1-19 ; Vasconcelos, N. (2007). *From pixels to semantic spaces: Advances in content-based image retrieval*, *Computer*, 40(7), 20-26.

49) Zhao, W., Chellappa, R., Phillips, P., & Rosenfeld, A. (2003). *Face recognition: A literature survey*, *ACM Computing Surveys*, 35(4), 399-458.

50) Google. (15 août 2006). *A better way to organize photos?* Consulté le 10 juillet 2007 sur : <http://googleblog.blogspot.com/2006/08/better-way-to-organize-photos.html>

51) Voir <http://www.riya.com/>

52) Ribeiro, J. (19 janvier 2006). *Riya photo search site readies for launch*. Consulté le 11 juillet 2007 sur : <http://www.pcworld.com/article/id,124407-c,searchengines/article.html>

53) Arrington, M. (15 juin 2006). *Riya 2.0 on the way; major strategy shift*. Consulté le 11 juillet 2007 sur : <http://www.techcrunch.com/2006/06/15/riya-20-on-the-way-major-strategy-shift/>

54) Arrington, M. (15 juin 2006). *Op.cit.*

55) Voir <http://www.polarrose.com/>. A la date de rédaction de cet article, Polar Rose se présente encore sous forme de beta-test à accès privé.

images apparaissant sur la quasi-totalité des sites Web, ce qui permettra aux utilisateurs d'attribuer des tags aux visages qu'ils reconnaissent lorsqu'ils naviguent sur le Web⁵⁶. Ces données seront ensuite envoyées au serveur central de Polar Rose ; quiconque visionnera une image contenant le visage en question pourra alors indiquer de qui il s'agit, et confirmer ou corriger son tag. Ainsi, en voyant une photo représentant Bill Gates mais portant le tag « Steve Jobs », les utilisateurs pourront rendre Polar Rose « plus intelligent » en corrigeant l'erreur. Ils pourront également rechercher sur le Web d'autres photos montrant ce personnage, sur la base du profil biométrique unique créé pour son visage par Polar Rose.

En résumé, les services de recherche audiovisuelle ont connu de nombreuses avancées technologiques, notamment l'intégration des images créées par les utilisateurs dans des index interrogeables, l'attribution courante de « tags » à des fichiers image et le déploiement de systèmes de reconnaissance faciale s'appuyant sur la participation des utilisateurs. Grâce à ces innovations, les fournisseurs d'outils de recherche sur Internet ont été en mesure d'étendre leur *portée* pour tenir compte de la quantité croissante d'images créées par les utilisateurs disponibles sur Internet. Certains moteurs de recherche se concentrent d'ailleurs uniquement sur la mise à disposition de contenus audiovisuels ; c'est le cas du projet de recherche multimédia européen Quaero, dont l'objet est notamment de faciliter l'accès aux archives du patrimoine culturel audiovisuel et aux bibliothèques numériques⁵⁷. Dans le même temps, si ces nouveaux outils de recherche audiovisuelle permettent aux utilisateurs de chercher des visages, des sons, des vidéos et des lieux, ils offrent aussi la possibilité aux fournisseurs d'outils de recherche de mieux connaître les souhaits et les désirs de leurs utilisateurs, ce qui optimise leur *mémoire* des intentions des utilisateurs. Ainsi, le fait de savoir si un utilisateur cherche habituellement des vidéos de célébrités ou des images de monuments européens peut aider un moteur de recherche à fournir les résultats les plus pertinents pour la recherche sur les mots-clés « Paris Hilton ». Au total, le moteur de recherche parfait est de plus en plus en passe de devenir audiovisuel.

4. Moteur de recherche audiovisuel parfait et confidentialité

Dans la deuxième partie de cet article, nous avons abordé en détail les menaces particulières, en matière de protection de la vie privée, qui sont apparues du fait de l'obtention d'une portée et d'une mémoire accrues au fil de la quête du moteur de recherche parfait. A mesure que la dimension audiovisuelle est de plus en plus indissociable du moteur de recherche parfait, de nouvelles menaces sur la vie privée, peut-être plus fortes encore, font leur apparition. La partie qui suit va exposer en détail comment les tentatives d'enrichir le « moteur de recherche parfait » de composantes audiovisuelles intensifient encore les inquiétudes évoquées ci-dessus concernant la portée et la mémoire « parfaites » de cet outil « parfait ».

4.1. Portée parfaite de la recherche audiovisuelle et confidentialité

Comme on l'a évoqué plus haut, la portée parfaite des moteurs de recherche restreint le sentiment que pouvaient auparavant avoir les utilisateurs que leur vie privée était protégée du fait même qu'elle se trouvait « dans l'ombre » ; ces outils ont en effet rendu visibles des sites Web, des documents et d'autres fichiers en ligne qui étaient précédemment difficiles à localiser ou à atteindre. L'avènement de composantes audiovisuelles sophistiquées accélère cette désintégration de la « vie privée protégée par l'ombre », tandis que les utilisateurs ont de moins en moins de moyens de savoir si leurs photos en ligne sont accessibles ou identifiables. Sur les sites de partage d'images tels que Flickr, il est désormais facile pour tout internaute de télécharger des photos sur un serveur, de leur attribuer des tags et de les partager, qu'il s'agisse de photos de lui-même ou d'autres personnes. Avec l'omniprésence des caméras de téléphones portables, il nous est possible de saisir n'importe quel moment de notre vie, de l'« étiqueter » et de le mettre en ligne aux yeux de tous. Grâce à l'intégration des photos de Flickr dans les résultats généraux de recherche d'images sur Yahoo!, les utilisateurs n'ont plus besoin d'être membres de la communauté Flickr (ou même d'être au courant de son existence) pour accéder aux images stockées par le service. Résultat : un individu peut être pris en photo, la photo postée en ligne,

56) Simonite, T. (19 décembre 2006). *Face-hunting software will scour Web for targets*. Consulté le 11 juillet 2007 sur : <http://technology.newscientist.com/article/dn10828-facehunting-software-will-scour-web-for-targets.html>

57) Voir <http://www.quaero.fr/>

étiquetée de son nom et rendue accessible via le moteur de recherche Yahoo, sans que cette personne n'en soit jamais informée⁵⁸. A une époque où les particuliers sont fréquemment pris en photo (notamment lors de manifestations publiques), cette nouvelle possibilité d'attribuer des tags à ces images et de les mettre à disposition de millions d'internautes menace les normes existantes en termes de vie privée, puisque l'image de tout un chacun se trouve exposée à un public de masse.

La généralisation de métadonnées très précises tend à exacerber encore cette préoccupation. Les photos numériques font généralement appel à la norme d'échange de données pour les fichiers d'images « Exif », qui enregistre diverses métadonnées dans chaque image – notamment la date et l'heure de sa création, le réglage de l'appareil photo, un onglet permettant de visualiser l'image sur l'écran d'affichage de l'appareil ou dans un logiciel et, éventuellement, des informations concernant les droits d'auteur. Beaucoup d'images sont également géocodées : des tags servent d'indications de localisation et sont ajoutés soit manuellement par l'utilisateur, soit automatiquement via un récepteur GPS intégré à l'appareil photo lui-même. Ces métadonnées détaillées peuvent être visualisées (et souvent faire l'objet d'une recherche) au sein de sites de partage de photos en ligne, permettant aux utilisateurs de savoir précisément quand et où ont été prises certaines photos. Résultat, non seulement un nombre croissant de photos de particuliers est disponible en ligne à leur insu, mais en outre, il est possible d'accéder gratuitement à de nombreuses informations détaillées quant au lieu précis où se trouvait la personne lorsque la photo a été prise et l'heure de réalisation du cliché, ce qui réduit encore la part « d'ombre » de la vie privée de cette personne.

Avec l'apparition de puissants outils de reconnaissance faciale permettant d'identifier et de chercher des images en ligne, nous nous trouvons face à un autre cas où la portée considérable du moteur de recherche audiovisuel parfait menace la vie privée de son utilisateur⁵⁹. Ces outils ne s'en remettent plus aux utilisateurs pour identifier le visage d'une personne et lui attribuer un tag sur une photo dans Flickr ; les visages peuvent désormais facilement être localisés sur la base de leurs caractéristiques biométriques propres. Grâce à des services tels que Riya ou Polar Rose, un utilisateur pourrait télécharger sur le serveur des dizaines de photos personnelles pour « apprendre » au logiciel à reconnaître son visage, ce qui rendrait plus facile l'étiquetage de futures photos. Riya rendrait ce profil facial disponible aux autres utilisateurs souhaitant rechercher et identifier d'autres images. Si un deuxième utilisateur soumettait ensuite une photo quelconque trouvée en ligne, le logiciel Riya scannerait alors sa base de données à la recherche de visages correspondants et étiquetterait automatiquement tout visage reconnaissable au sein de la nouvelle image. Les utilisateurs pourraient également balayer le Web tout entier à la recherche des visages qu'ils ont appris à reconnaître à Riya. Auparavant, il était difficile, pour ce deuxième utilisateur, d'identifier la personne représentée sur la photo ; avec les outils de reconnaissance faciale tels que Riya et Polar Rose, la reconnaissance de visages inconnus se trouve grandement facilitée. En outre, Polar Rose simplifie également la recherche sur le Web d'autres images comprenant le visage de cette même personne. Ceci peut conduire à trouver des images sur des sites de réseaux sociaux, des pages personnelles, des sites d'employeurs, etc. Si, auparavant, ces images étaient visibles par tous, l'utilisateur lambda ignorait le plus souvent leur existence, et ne pouvait que difficilement les trouver. Avec ces nouveaux outils de reconnaissance faciale destinés aux sites de réseaux sociaux, les utilisateurs perdent la capacité de contrôler le contexte et la communauté dans lesquels certaines images sont visibles.

4.2. Mémoire parfaite de la recherche audiovisuelle et confidentialité

Parallèlement aux menaces sur la vie privée nées de l'extension de la portée de la recherche audiovisuelle, des menaces similaires se font jour dès lors que la *mémoire* du moteur de recherche parfait se trouve accrue par la présence de contenus audiovisuels. Tout comme les requêtes générales des utilisateurs sont pistées et enregistrées par les moteurs de recherche, afin de prédire leurs souhaits et leurs besoins, les mots-clés employés lors d'une recherche audiovisuelle – noms, lieux, pornographie,

58) Ainsi, une recherche sur les termes « Michael Zimmer » via le service de recherche d'images Yahoo Image Search révèle une photo de moi donnant une conférence, portant un « tag » à mon nom et postée sur Flickr – le tout, à mon insu.

59) Granick, J. (5 juillet 2005). *Face it: Privacy is endangered*. Consulté le 8 avril 2008 sur : <http://www.wired.com/politics/law/commentary/circuitcourt/2005/12/69771> ; Inman, M. (5 janvier 2007). *Face recognition for online photo searches sparks privacy fears*. Consulté le 5 avril 2008 sur : <http://news.nationalgeographic.com/news/2007/01/070105-photo-search.html>

etc. – seront enregistrés de la même façon et pourront potentiellement être associés au compte de chaque utilisateur. Outre le fait qu'elles peuvent être tout simplement embarrassantes, certaines requêtes éclairent les opinions sociales et politiques d'un individu, et peuvent même être parfois d'une légalité douteuse. Les organismes chargés de l'application de la loi commencent déjà à s'intéresser au suivi et à la collecte des activités de recherche généralistes⁶⁰, mais elles pourraient aussi porter leur attention sur les recherches de fichiers audiovisuels qui concernent des éléments sujets à controverse ou illégaux – pédopornographie, caricatures du prophète Mohammed, objets de l'époque nazie ou fichiers protégés au titre du droit d'auteur. En d'autres termes, le type de requête unique, rendu possible par l'ajout de contenus audiovisuels dans les services de recherche en ligne, menace de rendre la « base de données d'intentions » plus détaillée encore et potentiellement plus nuisible.

5. Conclusion

Dans un effort visant à rendre les contenus audiovisuels (y compris les images animées) aussi « accessibles et utiles à tous » que leurs cousins textuels, de nouveaux outils puissants sont venus enrichir l'offre de services de recherche sur le Web : inclusion d'images créées par les utilisateurs, attribution exhaustive de tags, création de métadonnées, et apparition d'outils sophistiqués de reconnaissance faciale placés entre les mains des utilisateurs de la Toile au quotidien. Ces nouvelles améliorations dans le domaine audiovisuel suscitent de grandes inquiétudes quant à la protection des données privées ; en effet, la portée et la mémoire du moteur de recherche parfait permettent désormais d'y inclure les visages de particuliers, des identifiants biométriques, des métadonnées détaillées et l'historique des types de contenus multimédias recherchés par ces individus. C'est en quelque sorte un marché avec le diable qui voit le jour : les technologies de recherche audiovisuelle promettent d'améliorer encore le moteur de recherche parfait, en rendant disponible une vaste quantité de contenus multimédias – dont beaucoup sont créés par les utilisateurs – tout en s'appuyant aussi sur les métadonnées et la philosophie du Web 2.0, qui permet aux utilisateurs d'identifier et d'étiqueter des images sur Internet pour faciliter leur localisation future. Mais de nouvelles préoccupations voient aussi le jour en matière de confidentialité, notamment face à la disparition progressive de la protection de la vie privée « par l'ombre », au fait que les utilisateurs perdent la capacité de contrôler si des photos d'eux-mêmes sont téléchargées vers des serveurs ou si leur visage est lié à un tag, et enfin au fait que les activités de recherche audiovisuelle viennent s'ajouter à la « base de données d'intentions » toujours croissante constituée par les fournisseurs de moteurs de recherche.

Quelles possibilités s'offrent à nous pour renégocier ce marché « faustien » passé avec ces nouveaux outils de recherche audiovisuelle ? Tout d'abord, nous devons étendre notre conception de la protection de la vie privée en rapport avec les contenus disponibles en ligne, y compris les images mises en ligne par des particuliers ou les photos d'individus réalisées dans des lieux publics. Il devient essentiel d'adopter une conception plus contextuelle de la vie privée⁶¹ et d'admettre que le simple fait qu'un individu se trouve dans un lieu public n'autorise pas à prendre une photo de lui, à la télécharger vers le Web, à l'identifier, à lui attribuer des tags et éventuellement à réaliser un scanner biométrique de son visage qui sera partagé avec tous les internautes. Certaines images sont réservées à un contexte particulier, et ces nouvelles pratiques – même si elles sont motivées par une intention positive d'apporter des améliorations au moteur de recherche parfait – peuvent constituer des atteintes à la vie privée.

Ensuite, nous devons envisager de nouveaux cadres législatifs et réglementaires pour faire face au flot croissant d'images personnelles en ligne et à la montée en puissance de technologies sophistiquées visant à les identifier, leur attribuer des tags et permettre leur recherche sur la Toile. Certaines juridictions ont pris des mesures en vue d'interdire purement et simplement la publication de photos de personnes en ligne sans leur consentement explicite. D'autres exigent le floutage de tout visage identifiable dès lors que les individus représentés n'ont pas donné leur accord. On peut également

60) McCullagh, D. (20 décembre 2006). *Police blotter : Google searches nab wireless hacker*. Consulté le 13 juin 2007 sur http://www.news.com/2100-1030_3-6144962.html

61) Nissenbaum, H. (2004). *Privacy as contextual integrity*. *Washington Law Review*, 79(1), 119-157 ; Barth, A., Datta, A., Mitchell, J. C., & Nissenbaum, H. (2006). *Privacy and contextual integrity: Framework and applications*. Article présenté lors du symposium de l'IEEE sur la sécurité et la vie privée ; Zimmer, M. (2008). *Privacy on planet Google*. *Op.cit.*

prévoir la mise en place d'un système de « notification et retrait » comparable à celui qui est utilisé lors des allégations de violation du droit d'auteur selon le *Digital Millennium Copyright Act* américain. Dans ce cas, les utilisateurs rencontrant une photo d'eux-mêmes en ligne pourront adresser une notification à la personne qui a mis l'image en ligne ou le site qui l'héberge, afin que la photo en question soit retirée.

Enfin, diverses solutions technologiques seraient susceptibles d'atténuer les inquiétudes en matière de vie privée qui frappent les nouveaux outils de recherche audiovisuelle. Même en l'absence de contraintes législatives ou réglementaires, les fournisseurs d'outils de recherche audiovisuelle pourraient fournir les outils technologiques permettant aux individus de demander le retrait de leur photo s'ils la trouvent sur Internet, ou créer un système de signalement conséquent (comme dans le secteur du crédit) pour informer les particuliers dès lors qu'une photo de leur visage est « taggée » ou mise en ligne. Concernant le pistage et la collecte des recherches audiovisuelles des internautes dans des « bases de données d'intentions », les fournisseurs d'outils de recherche devraient permettre aux utilisateurs d'accéder aux activités ainsi collectées, d'en retirer toute information dont ils ne souhaitent pas qu'elle soit conservée, et d'interdire l'enregistrement de leurs futures requêtes.

En résumé, l'avènement de puissants outils de recherche audiovisuelle peut potentiellement servir la quête du moteur de recherche parfait. Mais il nous faut affronter les nombreuses inquiétudes quant à la protection de la vie privée sont traitées (en élargissant notre conception de la vie privée, en créant des protections relevant du politique, et en élaborant des systèmes de protection de la vie privée) afin d'éviter de suivre les pas de Faust et de conclure un marché avec le diable en échange d'un accès au savoir universel.

L'incidence de la liberté d'expression sur la gouvernance des outils de recherche

Joris van Hoboken

Institut du droit de l'information (IViR), Université d'Amsterdam

Introduction

Services et outils de recherche ont acquis une place centrale dans notre environnement informationnel. L'explication la plus simple de ce phénomène est l'abondance de l'offre d'informations en ligne, notamment sur le Web ; mais ce sont des innovations dans les technologies de la recherche et de l'information qui ont rendu possible cette évolution. Face à cette abondance, la question n'est désormais plus de savoir si telle ou telle information est publiée, mais si un utilisateur final sera capable de la trouver parmi toutes les autres données disponibles. Si Internet et les médias à base de texte se distinguent, c'est en raison de la quantité écrasante de sources et de contenus qu'ils offrent à leurs utilisateurs ; toutefois, les contenus audiovisuels les talonnent. Selon les projections statistiques de fréquentation d'Internet réalisées par Cisco, la « Vidéo à la demande (VoD), la télévision sur IP (IPTV) et la télévision sur Internet représenteront en 2012 près de 90 % du trafic IP total¹ ». Par conséquent, le pouvoir de « contrôleur d'accès » tend à se reporter sur de nouveaux prestataires intermédiaires, tandis que les questions liées à la gouvernance des outils et services d'indexation des contenus disponibles, et d'aide aux utilisateurs dans la sélection du contenu de leur choix, commencent à prendre de l'importance dans le débat sur les politiques à adopter dans la société de l'information.

Cet article entend éclairer le débat sur la réglementation des outils de recherche du point de vue de la liberté d'expression et d'information. Pour ce faire, nous étudierons certains aspects de la gouvernance des moteurs de recherche, du droit à la liberté d'expression aux intérêts des acteurs concernés au premier chef par la libre circulation d'informations qui se trouve facilitée par ces outils de recherche. Le cas échéant, nous nous intéresserons aux spécificités de la réglementation à appliquer à la recherche sur les médias audiovisuels.

Dans une première partie, nous décrirons les moteurs de recherche d'un point de vue fonctionnel et donnerons quelques exemples illustrant leur importance croissante au titre de maillon dans la chaîne de valeur des contenus Internet (1.1). Nous nous intéresserons ensuite aux acteurs concernés au premier chef par la gouvernance des moteurs de recherche (1.2) et à leurs intérêts dans le fonctionnement des moteurs de recherche (1.3).

Dans une deuxième partie, nous verrons dans quelle mesure ces intérêts sont protégés par le droit à la liberté d'expression et d'information, consacré par l'article 10 de la Convention de sauvegarde des droits de l'homme et des libertés fondamentales. Après un bref examen de cette question générale

1) *Cisco Visual Networking Index - Forecast and Methodology* (Prévision et méthodologie), 2007-2012, livre blanc, 16 juin 2008. Disponible sur : http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360.pdf, consulté le 1^{er} juillet 2008.

(2.1), notre analyse consistera à comparer les moteurs de recherche aux filtres Internet (2.2) et à étudier les questions qu'ils soulèvent en matière de liberté d'expression. Dans ce cadre, nous aborderons la récente Recommandation du Comité des ministres du Conseil de l'Europe sur les mesures visant à promouvoir le respect de la liberté d'expression et d'information au regard des filtres Internet (2.3)². Comme nous le verrons, la recommandation du Conseil de l'Europe sur le filtrage Internet, ainsi que le rapport³ qui la sous-tend, constituent un point de départ intéressant pour analyser l'incidence de la liberté d'expression sur la gouvernance de l'inclusion et de l'exclusion des services d'information par les moteurs de recherche ; en d'autres termes, en ce qui concerne la gamme de fournisseurs d'information répertoriés par le moteur de recherche à laquelle l'on peut accéder via un service de recherche particulier (2.4)⁴.

La troisième partie ouvre le débat quant à la responsabilité des moteurs de recherche en matière de contenus provenant de tiers, une question importante pour gérer l'inclusion et l'exclusion des fournisseurs d'information dans une plateforme de recherche. Cette partie présentera brièvement le régime juridique actuellement en vigueur au niveau de l'Union européenne (ci-après « UE ») en matière de responsabilité des moteurs de recherche et définira le point de vue de la liberté d'expression dans ce contexte. La quatrième partie forme la conclusion de l'article.

1. Les moteurs de recherche

1.1. A quoi sert un moteur de recherche ?

D'une manière générale, les moteurs de recherche aident les utilisateurs finaux d'un média donné à trouver des contenus (notamment audiovisuels) consultables grâce à ce média. Les moteurs de recherche en ligne, tels que Google et Yahoo! proposent ce type de services aux utilisateurs finaux d'Internet. Les acteurs primaires de ce service de recherche sont le prestataire lui-même, les utilisateurs finaux et les fournisseurs d'informations. Les moteurs de recherche sont actifs dans le secteur de l'organisation d'information. Les fournisseurs de moteurs de recherche à vocation commerciale réalisent généralement des bénéfices en vendant des espaces publicitaires ciblés parallèlement à leurs services. Ces publicités sont souvent liées aux recherches effectuées par les utilisateurs finaux. La plupart des moteurs de recherche sont à vocation commerciale et suivent un modèle économique de ce type. Pour les utilisateurs finaux, le moteur de recherche est un outil qui leur permet de sélectionner des informations à partir d'une liste mise à disposition via un média particulier. Derrière ce service de recherche, on trouve les fournisseurs d'information, qui proposent une offre d'informations susceptible d'être incluse dans le moteur de recherche. Pour eux, les moteurs de recherche représentent donc un moyen d'atteindre leur public ; la plateforme de recherche fonctionne comme un forum au sein duquel ils sont en concurrence pour être bien placés et recueillir l'attention des utilisateurs finaux. Il est donc important de figurer dans l'index d'un moteur de recherche, ce que l'on peut exprimer par la formule suivante : « Pour exister, il faut être répertorié par un moteur de recherche »⁵. Si la plupart des fournisseurs d'informations ont pour objectif d'y figurer et d'y apparaître en bonne place, certains préfèrent dans les faits être exclus des moteurs de recherche⁶.

La mise en œuvre d'un moteur de recherche passe par des étapes fonctionnelles que l'on peut regrouper en deux temps. D'abord, le prestataire doit composer une liste d'informations – ou index – à partir de la totalité des informations pouvant être mises à la disposition de l'utilisateur final.

2) Recommandation CM/Rec(2008)6 du Comité des ministres aux Etats membres sur les mesures visant à promouvoir le respect de la liberté d'expression et d'information au regard des filtres Internet. Disponible sur : [https://wcd.coe.int/ViewDoc.jsp?Ref=CM/Rec\(2008\)6](https://wcd.coe.int/ViewDoc.jsp?Ref=CM/Rec(2008)6), consulté le 1^{er} juillet 2008.

3) Rapport du Groupe de spécialistes sur les droits de l'homme dans la société de l'information (MC S IS) sur l'utilisation et l'impact des mesures de filtrage technique pour divers types de contenus dans l'environnement numérique, tel que publié dans le document CM(2008)37 add, disponible sur : [https://wcd.coe.int/ViewDoc.jsp?Ref=CM\(2008\)37&Ver=add](https://wcd.coe.int/ViewDoc.jsp?Ref=CM(2008)37&Ver=add), consulté le 1^{er} juillet 2008.

4) Schulz, W., T. Held, et al. (2005). *Suchmaschinen als Gatekeeper in der öffentliche Kommunikation Rechtliche Anforderungen an Zugangsoffenheit und Transparenz bei Suchmaschinen im WWW*.

5) Introna, L. D. et H. Nissenbaum (2000). *Shaping the Web: Why the Politics of Search Engines Matters*. *The Information Society* 16(3) : 169-186.

6) Voir *infra*, paragraphe 2.4.

Ensuite, il peut utiliser cette liste pour proposer un service de sélection interactif, qui permettra à l'utilisateur final d'effectuer une recherche dans cet index. En conséquence, un moteur de recherche en ligne permet aux utilisateurs finaux d'explorer son index, mais pas la totalité de la toile.

Dans le cas du moteur de recherche proposé sur Internet par Google, par exemple, les informations disponibles reflètent toutes les données en ligne auxquelles ses robots indexeurs (« spiders » et « crawlers ») peuvent accéder. Le moteur de recherche envoie ces robots à la recherche d'informations disponibles et génère automatiquement un index sur la base de ces résultats. Avec les nouvelles technologies, il est désormais possible de détecter des contenus audiovisuels, et un nombre croissant de moteurs de recherche ont élaboré des moyens de répertorier les contenus audiovisuels. Cependant, beaucoup d'informations présentes sur le Web ne sont pas incluses dans cet index. De nombreux fournisseurs d'information utilisent en effet le fichier d'instructions *robots.txt* conjugué à d'autres fonctionnalités, afin d'enjoindre aux robots indexeurs de ne pas répertorier certaines parties de leurs sites Web ; les grands moteurs de recherche respectent généralement ces instructions. Certaines sources peuvent ne pas être répertoriées pour d'autres raisons – par exemple, parce que les robots ne sont pas capables, techniquement, d'extraire les informations contenues sur un site particulier. Enfin, Google peut également ne pas inclure un site précis dans son index parce qu'il serait contraire à la loi d'afficher cette référence – par exemple parce qu'elle lui a été signalée au titre des dispositions de la *Digital Millennium Copyright Act* (DMCA). Par conséquent, l'index de Google contient une sélection d'informations disponibles sur le Web, sélection déterminée par les fournisseurs d'informations, par des tiers, par Google, par les propriétés techniques du Web et par la loi.

Une fois l'index réalisé, les utilisateurs finaux peuvent interagir avec le moteur de recherche en saisissant des mots-clés. Les résultats de la recherche sont organisés et hiérarchisés pour aider l'utilisateur à trouver des résultats satisfaisants ; en outre, des publicités ciblées apparaissent à côté des résultats, et permettent de générer des bénéfices. Les méthodes utilisées par les grands moteurs de recherche tels que Google pour organiser et hiérarchiser les contenus consistent en une combinaison complexe de méthodes statistiques. L'une d'elles est l'analyse de fréquence de mots-clés sur les sites Web. De nombreux moteurs de recherche effectuent également une analyse de la popularité et de l'importance des sites Web, en réalisant par exemple une analyse sur les hyperliens. PageRank, l'algorithme original de Google, est l'un des algorithmes les plus connus et les plus efficaces qui exploitent les hyperliens présents sur le Web pour juger de l'intérêt d'un site donné. Il mesure la pertinence d'un site (nommons-le « X ») en recensant le nombre de liens sur d'autres sites qui mènent vers « X ». Un lien a plus de poids s'il provient d'un site qui fait lui-même l'objet d'un grand nombre de liens. Par conséquent, un site est d'autant plus pertinent, aux yeux de PageRank, qu'il fait l'objet de nombreux liens provenant de pages elles-mêmes jugées fortement dignes d'intérêt. Les moteurs de recherche utilisent un autre facteur pour déterminer le classement d'un site Web : les consultations passées et le comportement de recherche de l'utilisateur final (des utilisateurs finaux). Cet historique de recherche et de consultation peut être utilisé pour personnaliser les résultats de la recherche et prodiguer des recommandations aux utilisateurs finaux. Le détail des algorithmes plus complexes de hiérarchisation des résultats est généralement gardé secret.

1.2. Exemples de réglementation sur les moteurs de recherche en Europe

Beaucoup ont pris conscience de l'importance des moteurs de recherche et de leur impact sur la circulation des données dans notre environnement informationnel. Leader du marché de la recherche sur Internet, Google attire tous les regards, mais les moteurs de recherche dans leur ensemble ont récemment pris de l'importance aux yeux des instances de réglementation et des décideurs européens.

Dans le cadre d'une réglementation sur les contenus en ligne, l'Ofcom (l'organisme de réglementation britannique, dont le champ de compétences couvre la télévision, la radio et les communications électroniques) a récemment désigné les moteurs de recherche comme un segment à part entière dans la chaîne de valeur Internet⁷. Il divise cette chaîne de valeur en deux grandes catégories, à savoir d'une part les entités qui contrôlent la mise à disposition des contenus (tels que les producteurs et agréga-

7) Cf. Ofcom - *Office of Communications*, réaction de l'Ofcom au rapport Byron, 27 mars 2008, p. 50-51. Disponible sur : http://www.ofcom.org.uk/research/telecoms/reports/byron/byron_review.pdf, consulté le 1^{er} juillet 2008.

teurs de contenus, ainsi que les hébergeurs) et, d'autre part, les entités contrôlant l'accès aux contenus (tels que les fournisseurs d'accès à Internet, les moteurs de navigation et de recherche, mais aussi les terminaux des clients). Selon l'Ofcom, les interventions et stratégies traditionnelles de réglementation en matière de diffusion des contenus ont perdu de leur efficacité en raison du caractère mondial du marché des contenus sur Internet et des complications liées à la compétence des juridictions⁸. Là où la réglementation des contenus concernait auparavant les entités qui, au sein de la chaîne de valeur, contrôlaient la mise à disposition des contenus, l'Ofcom souligne désormais la nécessité de s'intéresser aussi aux organismes qui contrôlent l'accès aux contenus. Il conclut que, dans une politique qui vise à gérer les risques liés aux contenus dangereux pour les enfants, l'autorégulation est un élément capital – notamment celle qu'opèrent les Fournisseurs d'Accès à Internet, qui ont un lien fort avec la juridiction dont dépendent les utilisateurs finaux. Dans le cas des moteurs de recherche, toutefois, on ne manquera pas de noter l'absence de tels liens avec les juridictions. L'Ofcom en conclut qu'il faut encourager le filtrage, qui est une solution possible, mais qu'on ne peut exiger légalement des FAI d'opérer un filtrage au niveau du réseau – un mode de contrôle des contenus, pratiqué par exemple en Chine et en Arabie Saoudite⁹.

Le débat européen sur la réglementation des moteurs de recherche s'intéresse aussi aux conséquences de leur utilisation en matière de vie privée et de protection des données personnelles. Le Groupe de travail « Article 29 » sur la protection des données, mis en place par l'Union européenne, a enquêté sur la politique de conservation des données des grands moteurs de recherche sur Internet¹⁰. Ces derniers tendent à conserver relativement longtemps les données relatives à l'utilisation de leurs services par les utilisateurs finaux. L'historique des moteurs de recherche comprend généralement les mots clés saisis, l'adresse IP (les adresses IP) à partir de laquelle (desquelles) l'utilisateur se connecte à Internet tel jour à telle heure, ainsi qu'un cookie unique identifiant l'utilisateur et sa navigation sur la plateforme du moteur de recherche une fois que les résultats de sa recherche lui ont été présentés. Le Groupe de travail « Article 29 » a récemment achevé une enquête sur les pratiques des grands moteurs de recherche en matière de traitement des données, et publié un avis officiel détaillé qui s'ouvre sur l'affirmation suivante : « Dans la société de l'information, les fournisseurs de moteurs de recherche sur Internet jouent un rôle essentiel d'intermédiaires. Le groupe de travail reconnaît la nécessité et l'utilité des moteurs de recherche, et il est conscient de leur contribution au développement de la société de l'information¹¹ ».

Enfin, il est également question des moteurs de recherche dans le rapport du Conseil de l'Europe consacré aux conséquences de la liberté d'expression pour les filtres de contenus sur Internet¹². Il précise : « Aujourd'hui, la plupart des internautes utilisent un moteur de recherche lorsqu'ils ont besoin de rechercher rapidement les informations qui satisferont leurs besoins. Le moteur de recherche joue donc un rôle de contrôleur d'accès du fait qu'il sélectionne les informations parvenant à l'utilisateur et détermine la forme sous laquelle elles lui parviennent ; un rôle d'autant plus important que, comme c'est le cas actuellement, le marché des moteurs de recherche est dominé par un seul fournisseur. La façon dont ce moteur de recherche affiche les résultats et l'ordre de priorité dans lequel ils apparaissent influence le type et la nature des informations reçues par l'utilisateur¹³. » Dans la deuxième partie de cet article, nous étudierons plus avant les conséquences de la liberté d'expression pour les moteurs de recherche et les filtres de contenus en ligne.

8) *Ibid.*, p. 50-51.

9) *Ibid.*, p. 51.

10) Le Groupe de travail « Article 29 » sur la protection des données est constitué des organismes nationaux de protection des données des Etats membres de l'UE ; il a pour mission de faciliter l'harmonisation de l'interprétation de la Directive sur la protection des données (95/48/CE) et de la Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, JO L 281, 23 novembre 1995, p. 31-50.

11) Groupe de travail « Article 29 », avis sur les aspects de la protection des données liés aux moteurs de recherche, WP 148, 4 avril 2008. Dans plusieurs passages de l'avis, le groupe de travail aborde l'importance du droit fondamental à la liberté d'expression et d'information dans le contexte des moteurs de recherche. (« [...] Il faut parvenir à un équilibre, dans le droit communautaire en matière de protection des données et les droits des divers Etats membres, entre la protection du droit à la vie privée et la protection des données à caractère personnel d'une part, et la libre circulation de l'information et le droit fondamental à la liberté d'expression, d'autre part. »). Disponible sur : http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_fr.pdf, consulté le 1^{er} juillet 2008.

12) Voir *supra*, notes 2 et 3 et texte correspondant.

13) *Supra*, note 3.

1.3. Intérêts des fournisseurs d'informations, des utilisateurs finaux, des moteurs de recherche et des tiers

Avant d'aborder certaines des incidences de la liberté d'expression sur la gouvernance de l'accessibilité des contenus via les moteurs de recherche, il est éclairant d'étudier les intérêts des acteurs concernés au premier chef par la question. Nous proposerons dans la partie qui suit une définition globale des enjeux que pose l'accessibilité des contenus via une plateforme de recherche pour le fournisseur de moteurs de recherche, pour les utilisateurs finaux et pour les fournisseurs d'information. Comme on le verra, leurs intérêts ne coïncident pas toujours. Un peu plus loin dans cet article, nous nous demanderons dans quelle mesure ces intérêts sont protégés au regard de la liberté d'expression. Bien entendu, des parties tierces peuvent également avoir des intérêts dans la gouvernance des moteurs de recherche – tels que la protection de la vie privée, de la réputation ou de la propriété intellectuelle. Ces aspects dépassent toutefois le champ du présent article¹⁴.

Si l'on commence par le processus de recherche en lui-même, différents types de questions se posent quant à la gouvernance de l'accessibilité des contenus. Comme nous l'avons expliqué plus haut, la composition de l'index joue un rôle décisif dans les résultats que peut fournir un moteur de recherche donné, et l'exclusion d'une source particulière de l'index d'un moteur de recherche peut avoir des raisons variées. Toutefois, au-delà de la composition de l'index, la hiérarchisation des résultats de la recherche a également un impact considérable sur les informations qui pourront être obtenues. Des recherches empiriques montrent que les utilisateurs finaux ne dépassent généralement pas la première page de résultats et ont tendance à cliquer davantage sur les deux premiers résultats recensés : ainsi, il faut qu'un film figure sur la première page des résultats pour pouvoir être trouvé¹⁵. Pour exister, il ne suffit donc pas seulement d'être indexé. Un secteur entier spécialisé dans le marketing et l'optimisation des moteurs de recherche a par conséquent vu le jour pour aider les fournisseurs d'informations à figurer dans les premiers résultats des recherches et à gérer des campagnes de publicité efficaces au sein des moteurs de recherche. Pour finir, le service de recherche peut être personnalisé en fonction de publics et/ou d'utilisateurs finaux particuliers. Nous pourrions aboutir à une situation dans laquelle chaque utilisateur final disposerait d'un service de recherche personnalisé en fonction de ses centres d'intérêt et du contexte d'utilisation.

Tous les acteurs primaires ont ainsi leurs propres intérêts vis-à-vis de ces questions de gouvernance. Il est de l'intérêt du fournisseur de moteurs de recherche de contrôler son propre service, et notamment la composition de son index, son fonctionnement, ainsi que la pertinence et l'intégrité des résultats de la recherche. Il est de l'intérêt de nombreux fournisseurs d'informations d'être placés efficacement dans un moteur de recherche afin de toucher leur public, même s'il faut signaler que pour des raisons variées, certains ne souhaitent pas, en fait, figurer dans un moteur de recherche. Puisque les modèles économiques des moteurs de recherche et de nombreux éditeurs en ligne se fondent sur les revenus publicitaires, bien des éditeurs sont également en concurrence avec les moteurs de recherche pour percevoir des revenus de la publicité en ligne¹⁶. Enfin, il est de l'intérêt de l'utilisateur final, d'une manière générale, de pouvoir effectuer sa recherche librement et de trouver les fournisseurs d'informations qui offrent le service le plus avantageux. Il est donc de son intérêt de pouvoir s'informer de façon exhaustive en utilisant ce média. Dans la partie qui suit, nous aborderons le droit à la liberté d'expression et d'information au vu des intérêts de ces acteurs primaires.

14) Pour une présentation complète des intérêts en jeu dans la gouvernance des différents aspects des moteurs de recherche et de la diffusion des informations qu'ils facilitent, voir James Grimmelmann, *The Structure of Search Engine Law*, *Iowa Law Review*, vol. 93, n° 1, 2007. Disponible sur : <http://ssrn.com/abstract=979568>

15) A ce sujet et pour d'autres résultats empiriques concernant l'utilisation des moteurs de recherche, voir le numéro spécial du *Journal of Computer-Mediated Communications*, volume 12, n° 3, avril 2007, Special Theme I, *The Social, Political, Economic, and Cultural Dimensions of Search Engines*, publié par Eszter Hargittai. Disponible sur : <http://jcmc.indiana.edu/vol12/issue3/>, consulté le 1^{er} juillet 2008.

16) La Commission européenne aborde cette question dans sa récente communication sur les contenus créatifs en ligne. Cf. Communication de la commission au parlement européen, au conseil, au comité économique et social européen et au comité des régions sur les contenus créatifs en ligne dans le marché unique, COM(2007) 836 final, Bruxelles, 3 janvier 2008, p. 34-35.

2. Liberté d'expression et d'information, et moteurs de recherche

2.1. Les conséquences de l'article 10 de la Convention européenne des Droits de l'Homme

Aux termes de l'article 10 de la Convention européenne des Droits de l'Homme, toute personne a droit à la liberté d'expression ; ce droit comprend la liberté d'opinion et celle de recevoir ou de communiquer des informations ou des idées, sans qu'il puisse y avoir ingérence d'autorités publiques et sans considération de frontière¹⁷. L'article 10, paragraphe 2, dispose que ces libertés peuvent être soumises à d'éventuelles restrictions prévues par la loi, qui sont nécessaires dans une société démocratique et servent certains objectifs énumérés – tels que la protection de la santé ou de la morale, la protection de la réputation ou des droits d'autrui. La « protection des droits d'autrui » recouvre notamment la protection des droits de la propriété intellectuelle¹⁸.

La Cour européenne des droits de l'homme (ci-après « CEDH ») n'a à ce jour abordé la question des moteurs de recherche dans aucun de ses arrêts ; il est donc nécessaire d'étudier les autres formes de jurisprudence de cette instance pour évoquer les conséquences de l'article 10 en matière de gouvernance des moteurs de recherche. Parmi les jugements comportant des observations sur ce thème, citons notamment l'affaire *Autronic*, dans lequel la Cour européenne des droits de l'homme réaffirme que l'article 10 ne s'applique pas seulement aux personnes physiques, mais aussi aux entreprises à but lucratif. Elle conclut :

« ni le statut juridique de société anonyme, ni le caractère commercial de ses activités ni la nature même de la liberté d'expression ne sauraient priver *Autronic AG* du bénéfice de l'article 10. Ce dernier vaut pour 'toute personne', physique ou morale¹⁹. »

De ce fait, la nature commerciale ou à but lucratif des grands moteurs de recherche ne les prive pas d'une possible protection par l'article 10. Dans le même jugement, la cour précise que :

« [l'article 10] concerne non seulement le contenu des informations mais aussi les moyens de transmission ou de captage, car toute restriction apportée à ceux-ci touche le droit de recevoir et communiquer des informations²⁰ ».

Il est difficile de savoir si les moteurs de recherche en général doivent être considérés comme des moyens de transmission et de captage d'informations. L'article 10 n'évoque pas explicitement un droit à chercher ou à recueillir des informations, mais on considère généralement que ces processus de communication ne doivent pas être exclus de son champ d'application. Il est intéressant de noter que tant la Déclaration universelle des droits de l'homme des Nations-Unies que le Pacte international relatif aux droits civils et politiques des Nations-Unies font explicitement référence, dans leur article 19, au droit de chercher des informations et des idées²¹.

Nico van Eijk a déjà souligné les difficultés qu'il y a à caractériser les moteurs de recherche au regard de la liberté d'expression, en raison de leurs multiples rôles²². Parce qu'il est à la fois un intermédiaire

17) Conseil de l'Europe, Convention de sauvegarde des Droits de l'Homme et des Libertés fondamentales, 4 novembre 1950, STE n° 5.

18) Pour une présentation, cf. P. Bernt Hugenholtz, *Copyright and Freedom of Expression in Europe*, dans : R. Dreyfuss et al. (éd.), *Expanding the Boundaries of Intellectual Property. Innovation Policy for the Knowledge Society*, New York : Oxford University Press, 2001, p. 343-363.

19) CEDH, 22 mai 1990, A 178, § 47 (*Autronic AG contre Suisse*).

20) *Ibid.*

21) La Déclaration universelle des droits de l'homme des Nations-Unies, adoptée et proclamée par l'Assemblée générale dans sa résolution 217 A (III) du 10 décembre 1948 (article 19 : « Tout individu a droit à la liberté d'opinion et d'expression, ce qui implique le droit de ne pas être inquiété pour ses opinions et celui de chercher, de recevoir et de répandre, sans considérations de frontières, les informations et les idées par quelque moyen d'expression que ce soit. ») ; Pacte international des Nations-Unies relatif aux droits civils et politiques, adopté et ouvert à la signature, à la ratification et à l'adhésion par l'Assemblée générale dans sa résolution 2200 A (XXI) du 16 décembre 1966 (article 19, paragraphe 2 : « Toute personne a droit à la liberté d'expression ; ce droit comprend la liberté de rechercher, de recevoir et de répandre des informations et des idées de toute espèce, sans considération de frontières, sous une forme orale, écrite, imprimée ou artistique, ou par tout autre moyen de son choix. »).

22) N. A. N. M. van Eijk, « Moteurs de recherche : Cherche et tu trouveras ? Positionnement juridique des moteurs de recherche », IRIS *plus* 2006-2, p. 5. Disponible sur : http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf, consulté le 1^{er} juillet 2008.

entre les fournisseurs d'information et les utilisateurs finaux, et lui-même un communicateur, le moteur de recherche est un média complexe. Le fournisseur d'accès à Internet d'un utilisateur final est un élément essentiel de toute communication en ligne entre cet utilisateur final et les fournisseurs d'information en ligne. Si le fournisseur d'accès ne facilite pas la communication entre l'utilisateur final et le fournisseur d'information, celle-ci ne peut pas avoir lieu. A l'inverse, les moteurs de recherche ne sont pas un élément essentiel de cette chaîne de communication linéaire qui relie les fournisseurs d'information aux utilisateurs finaux. Il est facile d'accéder à un fournisseur d'informations en ligne, soit en tapant directement son URL dans un navigateur, soit indirectement, en suivant un lien affiché ailleurs (par exemple sur un autre site Web), ou en utilisant un autre service de recherche ou de recommandation.

C'est pour ces raisons que Van Eijk affirme que les libertés de recevoir ou de communiquer des informations, telles qu'elles sont mentionnées explicitement dans l'article 10, ne constituent pas l'objet principal des moteurs de recherche, puisque l'information est déjà présente et accessible directement²³. Un moteur de recherche n'y change rien et ne peut rien y changer. Van Eijk en conclut que les moteurs de recherche *facilitent l'accès* à l'information mais n'offrent pas par eux-mêmes un accès à l'information. Il affirme que cette activité qui consiste à *rendre accessible* l'information devrait avoir un statut similaire, au regard de l'article 10, que celle consistant à divulguer ou diffuser des informations et des idées.

Cette approche est centrée sur la phase de communication ultime entre les fournisseurs d'informations et les utilisateurs finaux, (partiellement) facilitée par les moteurs de recherche. Elle doit recevoir une attention particulière, en raison de l'importance croissante des moteurs de recherche qui, dans notre environnement informationnel, sont des médiateurs, des éclaireurs et des moyens d'atteindre un public. La question, toutefois, est de savoir si l'article 10 ne protège pas également les communications qui ont lieu entre le fournisseur d'informations et le moteur de recherche, ainsi que celles qui se déroulent entre le moteur de recherche et l'utilisateur final. En se focalisant sur les communications intervenant entre ces acteurs primaires, nous pourrions peut-être analyser plus en détail l'incidence de la liberté d'expression sur la gouvernance des moteurs de recherche.

Les robots indexeurs du moteur de recherche constituent des moyens de recevoir des informations – d'une certaine façon, ils sont une version moderne et automatisée du chercheur ou du journaliste, qui part en quête de sources d'informations. L'arrêt *Autronic* de la CEDH précise que l'article 10 protège également les moyens de transmission et de captage. Puisque cette recherche d'informations (« *crawling* ») est un moyen de captage, elle devrait être protégée aux termes de l'article 10, et toute ingérence des pouvoirs publics doit respecter les conditions énumérées dans l'article 10, paragraphe 2. Par conséquent, toute loi qui fixerait des restrictions au « *crawling* » opéré par les moteurs de recherche devrait répondre à ces conditions. En outre, l'importance des moteurs de recherche pour la diffusion de l'information et des idées, et pour la capacité d'information du grand public, pourrait constituer un élément décisif pour déterminer dans quelle mesure le « *crawling* » et l'indexation sont protégés par l'article 10, et si de telles ingérences sont conformes à son paragraphe 2. Ainsi, la protection contre les ingérences dans les opérations de « *crawling* » pourrait être renforcée en raison de l'importance des moteurs de recherche dans notre société de l'information.

De même, la publication des résultats d'une recherche par un moteur de recherche accessible au grand public, ainsi que la demande et la réception de résultats par l'utilisateur final, tomberaient sous la protection de l'article 10. La publication des résultats de la recherche en elle-même constitue une communication entre le fournisseur du moteur de recherche et les utilisateurs finaux. De ce point de vue, le moteur de recherche ne se contente pas de *faciliter l'accès à l'information* de tiers, mais il *offre* aussi *l'accès à une information* sur des tiers. Dans ce contexte, il est intéressant de se demander dans quelle mesure les communications entre un moteur de recherche et ses utilisateurs sont en réalité de nature publique. On peut soutenir que la demande saisie par les utilisateurs d'un moteur de recherche constitue une communication privée, et non publique. Toutefois, dans la mesure où le service de recherche est disponible publiquement, la communication des résultats de la recherche aux utilisateurs est de nature publique²⁴.

23) *Ibid.*

24) Cf. Schulz & Held 2005.

Sur la base de la jurisprudence de la CEDH concernant l'article 10 et l'interprétation des « devoirs et des responsabilités » évoqués dans son paragraphe 2, plusieurs points méritent d'être mentionnés dans le contexte des moteurs de recherche. Premièrement, l'impact potentiel du média, ainsi que la nature du contenu susceptible d'être trouvé grâce à un moteur de recherche pourraient avoir leur importance, ici. A plusieurs reprises, la cour a précisé que l'impact particulier du média audiovisuel doit être pris en compte lorsqu'il s'agit de juger de la recevabilité d'une ingérence des pouvoirs publics. Dans l'arrêt *Jersild*, par exemple, la cour a conclu :

« S'agissant des 'devoirs et responsabilités' d'un journaliste, l'impact potentiel du moyen concerné revêt de l'importance et l'on s'accorde à dire que les médias audiovisuels ont des effets souvent beaucoup plus immédiats et puissants que la presse écrite [...]. Par les images, les médias audiovisuels peuvent transmettre des messages que l'écrit n'est pas apte à faire passer²⁵. »

Contre ces ingérences, les moteurs de recherche audiovisuels pourraient donc bénéficier d'une protection moindre que ceux qui sont à base de texte.

Deuxièmement, le rôle d'un moteur de recherche dans l'environnement informationnel, et la nature (technique) du service qu'il offre, peuvent lui valoir des devoirs et des responsabilités particulières. Selon la CEDH, les devoirs et responsabilités liés à l'article 10 de la Convention européenne des Droits de l'Homme dépendent également des moyens techniques utilisés pour l'expression et la diffusion. La cour a affirmé que :

« [...] quiconque exerce sa liberté d'expression assume des 'devoirs et responsabilités' dont l'étendue dépend, en particulier, du procédé technique utilisé. L'impact potentiel dudit procédé doit être pris en considération dans l'examen de la proportionnalité de l'ingérence. La garantie que l'article 10 offre à la presse est subordonnée, en raison de ces mêmes 'devoirs et responsabilités', à la condition que celle-ci agisse de manière à fournir des informations dignes de crédit dans le respect de la déontologie journalistique²⁶. »

Elle a ajouté, dans un jugement plus récent :

« Ces considérations jouent un rôle particulièrement important de nos jours, vu le pouvoir qu'exercent les médias dans la société moderne, car non seulement ils informent, mais ils peuvent en même temps suggérer, par la façon de présenter les informations, comment les destinataires devraient les apprécier. Dans un monde dans lequel l'individu est confronté à un immense flux d'informations, circulant sur des supports traditionnels ou électroniques et impliquant un nombre d'auteurs toujours croissant, le contrôle du respect de la déontologie journalistique revêt une importance accrue²⁷. »

Ces dernières considérations semblent en effet trouver un écho dans le contexte des moteurs de recherche, mais il reste à savoir si elles peuvent y être transposées, et si oui, de quelle façon. On peut se demander en outre si l'avenir verra le développement d'un corpus de déontologie et d'autorégulation parmi les moteurs de recherche. Ces derniers ont sans aucun doute un impact considérable sur notre environnement informationnel. Il serait toutefois prématuré de porter un jugement définitif sur ces questions. La partie qui suit prolonge le débat en se focalisant sur un aspect particulier de la gouvernance des moteurs de recherche au regard de la liberté d'expression, à savoir la gouvernance de la composition de l'index des moteurs de recherche. Néanmoins, avant d'aborder cette question, il est utile d'étudier un problème similaire : la gouvernance des filtres de contenus Internet, qui a récemment été abordée dans une recommandation du Comité des ministres du Conseil de l'Europe.

2.2. Filtres Internet et moteurs de recherche : une possible analogie ?

Il existe de nombreux types de filtres de contenus en ligne (« filtres Internet ») et ils sont mis en œuvre dans des circonstances variées. Nous nous intéresserons ici au type de filtres dont l'objectif est de limiter l'accessibilité des contenus Internet. Nous laisserons de côté les aspects techniques

25) CEDH 23 septembre 1994, § 31 (*Jersild contre Danemark*).

26) CEDH 14 juin 2007, § 42 (*Hachette Filipacchi Associés contre France*).

27) CEDH 10 décembre 2007, § 104 (*Stoll contre Suisse*).

d'Internet, mais signalons tout de même qu'il existe des filtres capables, entre autres choses, de détecter des contenus audiovisuels (Google, par exemple, travaille sur un filtrage des contenus protégés par copyright sur YouTube, sur la base d'une technique d'« empreinte digitale »)²⁸. Un logiciel de filtrage Internet vise à bloquer certaines catégories de contenus qui seraient accessibles aux utilisateurs finaux via une connexion Internet libre et non filtrée. Les filtres de contenus Internet peuvent être utilisés et installés sur les terminaux Internet par les utilisateurs finaux. Ils peuvent être mis en œuvre par des acteurs privés sur leurs réseaux – par exemple par les employeurs ou les propriétaires de cafés Internet. Ils peuvent être appliqués délibérément par les Fournisseurs d'accès à Internet (ci-après « FAI ») au niveau du réseau, au titre de service à leurs clients, en offrant à ceux-ci un contrôle sur le filtrage. Ils peuvent être installés par les bibliothèques publiques, qui souhaitent restreindre l'accès à des contenus nuisibles et illégaux sur leurs points d'accès à Internet²⁹. Dans un nombre croissant de pays, en Europe et ailleurs, les FAI sont convenus avec les pouvoirs publics de filtrer délibérément la pédopornographie au niveau du réseau. Dans certaines juridictions, les pouvoirs publics exigent des FAI par la loi qu'ils utilisent des modes de filtrage au niveau du réseau³⁰.

La fonction des filtres de contenus Internet consiste à bloquer une partie des informations disponibles sur le Web qui seraient accessibles autrement. En tant que tels, ils appartiennent à la catégorie plus large des intermédiaires de l'extraction de données, une catégorie de prestataires qui sélectionnent et filtrent les informations circulant sur Internet, entre les fournisseurs d'informations et les utilisateurs finaux³¹. Parmi les intermédiaires de l'extraction de données, citons les filtres de contenus Internet, les moteurs de recherche, les services de recommandation et les FAI qui bloquent ou filtrent certaines pages sur la base de leur contenu. Ces intermédiaires remplissent une fonction importante dans notre environnement informationnel. Dans ce milieu caractérisé par l'abondance, les intermédiaires de l'extraction de données aident les utilisateurs finaux à trouver et à sélectionner les informations qu'ils jugent pertinentes ou utiles, et excluent les informations auxquelles ils ne souhaitent pas accéder ou auxquelles ils n'ont pas le droit d'accéder – par exemple parce qu'elles sont nuisibles ou illégales.

Les moteurs de recherche et les filtres Internet étant similaires – au sens où tous deux prennent des décisions quant à la gamme de fournisseurs d'informations en ligne qui sont accessibles via leurs services –, il est intéressant de voir ce qui peut être dit sur les filtres Internet du point de vue de la liberté d'expression. En tenant compte des différences entre filtres Internet et moteurs de recherche, cette réflexion constituera un point de départ utile pour l'analyse de l'incidence de la liberté d'expression sur les moteurs de recherche. Première différence, les moteurs de recherche sont plus sophistiqués, car ils ne se contentent pas de sélectionner des fournisseurs d'informations – c'est-à-dire les sources, sur Internet, qui peuvent être trouvées grâce à leur service – mais en outre, ils organisent et rangent cette sélection de fournisseurs d'informations grâce à leurs algorithmes de classement. Autre différence, les filtres de contenus Internet sont généralement placés entre l'utilisateur final et Internet, au niveau du terminal de l'utilisateur final ou à un niveau quelconque du réseau ; de plus, le passage d'un filtre Internet à un autre est différent du passage d'un moteur de recherche à un autre. Bien qu'un moteur de recherche soit aujourd'hui dominant en termes d'utilisation sur le Web (Google), il en existe d'autres qui sont comparables, et les utilisateurs finaux peuvent normalement passer facilement de l'un à l'autre. Pour finir, l'objet des filtres de contenu Internet est d'empêcher l'accès aux contenus, tandis que celui des moteurs de recherche est d'aider les utilisateurs à accéder aux contenus. D'un point de vue fonctionnel, ces deux visées sont plutôt en opposition.

Toutefois, si l'on met à part ces différences, un parallèle important demeure entre la gouvernance des filtres Internet et des moteurs de recherche, en particulier lorsque l'on restreint le champ à la gouvernance de l'inclusion et de l'exclusion des fournisseurs d'informations dans l'index d'un moteur

28) Pour une présentation, voir le rapport du Groupe de spécialistes sur les droits de l'homme dans la société de l'information, ainsi que les références *supra*, note 3.

29) Aux Etats-Unis, certains financements gouvernementaux aux bibliothèques publiques ont été soumis à la condition que de tels logiciels de filtrage soient installés. Cf. Cour suprême des Etats-Unis, *United States v. American Library Association*, 539 U.S. 194 (2003).

30) Pour une présentation complète et les enjeux du filtrage mondial d'Internet, cf. Ronald Deibert, John Palfrey et al. (éd.), *Access Denied: The Practice and Policy of Global Internet Filtering*, Cambridge MA : The MIT Press, 2007.

31) Cf. Chandler, Jennifer A., *A Right to Reach an Audience: An Approach to Intermediary Bias on the Internet*, *Hofstra Law Review*, 2008. Disponible sur : <http://ssrn.com/abstract=1021344>

de recherche. Comme on l'a signalé plus tôt, ces deux types de services gèrent l'accessibilité de l'information. Pour un fournisseur d'informations, se voir ou non indexé par les moteurs de recherche, de même que figurer ou non sur la liste de blocage d'un filtre, ont des effets comparables sur son accessibilité au grand public. Pour les utilisateurs finaux, la composition de l'index des moteurs de recherche présente des enjeux similaires d'accès à l'information, de transparence et de contrôle, que les filtres de contenus Internet.

Il est intéressant de noter que les moteurs de recherche ont déjà recours à des filtres dans le cadre de leurs services. La fonction « recherche d'images » des grands moteurs de recherche dispose systématiquement d'un filtre optionnel que l'utilisateur final peut activer ou désactiver. Google offre ainsi l'option « *SafeSearch* » parmi les « préférences » applicables aux recherches, ce qui permet aux utilisateurs finaux d'opter pour un filtrage des images à caractère sexuel parmi les résultats d'une recherche d'images, et pour un filtrage de tous les résultats de la recherche, y compris lorsque cette dernière porte sur du texte³². Les préférences de filtrage sont sauvegardées grâce à l'installation d'un cookie dans le navigateur de l'utilisateur final. Google commence également à offrir un filtrage à ses entreprises clientes³³. Par suite des problèmes de spam touchant les moteurs de recherche (« *spamdexing* »), ces derniers peuvent également filtrer les pages qui ont pour seul but de manipuler les résultats des recherches et d'apparaître en bonne place de façon induue. Dans ce contexte, l'une des questions récurrentes est de savoir si les moteurs de recherche devraient avoir recours à des filtres pour empêcher l'affichage de ces résultats illégitimes, voire, être tenus de les utiliser, par la législation ou la réglementation.

Ces considérations sont en accord avec les conclusions du Groupe de spécialistes du Conseil de l'Europe sur les droits de l'homme dans la société de l'information, qui a élaboré un rapport sur les filtres Internet à destination du Comité des ministres³⁴. Le rapport souligne : « Les effets filtrants des moteurs de recherche sont [...] plus indirects que ceux des produits de filtrage », « [...] les produits de filtrage visent à protéger l'utilisateur contre des contenus illicites, préjudiciables ou inappropriés, les moteurs de recherche visent à fournir à l'utilisateur un accès rapide aux informations les plus susceptibles de correspondre à ses souhaits tels qu'exprimés par les termes de recherche ». Le rapport remarque aussi l'usage des filtres de contenus par les moteurs de recherche et affirme que « les conclusions établies ici valent pour les fournisseurs de moteurs de recherche dans la mesure où ils sont également des utilisateurs de mesures de filtrage ». Il déclare enfin « qu'au moins certaines lignes directrices identifiées dans ce rapport pourraient probablement servir pour des lignes directrices similaires concernant les moteurs de recherche³⁵ ».

Il ressort clairement de ce qui précède que, malgré l'existence de différences importantes, une étude des filtres Internet et de l'incidence qu'a la liberté d'expression sur eux est pertinente. Cette étude se fera dans l'esprit de la récente recommandation sur les filtres Internet et la liberté d'expression du Comité des ministres du Conseil de l'Europe.

2.3. Recommandation sur les filtres Internet et la liberté d'expression

Le 26 mars 2008, le Comité des ministres du Conseil de l'Europe a adopté une Recommandation CM/Rec(2008)6 sur les mesures visant à promouvoir le respect de la liberté d'expression et d'information au regard des filtres Internet³⁶. Cette recommandation et le rapport qui la sous-tend reconnaissent à la fois le rôle légitime des filtres Internet, et les façons dont ces filtres peuvent peser sur la liberté d'expression et d'information. Elle aborde explicitement certaines exigences posées par l'article 10 de la Convention européenne des Droits de l'Homme dans ce contexte, dont certaines concernent les intérêts des fournisseurs d'informations et des utilisateurs finaux. La recommandation en appelle aux Etats parties à la Convention européenne des Droits de l'Homme pour prendre des mesures concernant les filtres Internet, et propose dans cette optique une série de lignes directrices encourageant

32) *Google SafeSearch filtering preferences*, consulté le 1^{er} juillet 2008, disponible sur :

<http://www.google.com/support/bin/static.py?page=searchguides.html&ctx=preferences>

33) *Web Security for Enterprise*, Cnet News.com, 8 mai 2008. Disponible sur :

http://www.news.com/8301-10784_3-9938812-7.html?tag=nefd.top, consulté le 1^{er} juillet 2008.

34) Rapport du Groupe de spécialistes sur les droits de l'homme dans la société de l'information, et références, voir *supra*, note 3.

35) *Ibid.*

36) *Supra*, note 2.

l'information de l'utilisateur et sa sensibilisation, le contrôle des filtres Internet par les utilisateurs, ainsi que la responsabilité des acteurs privés et publics. Ces lignes directrices sont classées en trois parties. La première recense des lignes directrices pour l'« Utilisation et [le] contrôle des filtres Internet pour exercer et jouir pleinement de la liberté d'expression et d'information ». Dans la deuxième, il est question de la « Mise en place d'un filtrage approprié pour les enfants et les jeunes » et dans la troisième, de l'« Utilisation et [de la] mise en œuvre de filtres Internet par les secteurs public et privé ».

La recommandation opère une distinction entre, d'une part, le filtrage obligatoire et l'utilisation de filtres Internet par des organismes publics tels que les bibliothèques publiques et les écoles et, d'autre part, leur emploi par des entités privées telles que les entreprises dans le contexte de l'accès à Internet sur le lieu de travail. Le texte ne se limite pas à l'incidence de la liberté d'expression sur les filtres Internet dans le secteur public ou sur le filtrage prescrit par la loi, mais il comprend aussi des lignes directrices spécifiques quant à l'utilisation des filtres dans le secteur privé. Concernant l'utilisation et la mise en œuvre des filtres Internet par les organismes publics, ou le filtrage obligatoire, la recommandation qualifie l'utilisation de filtres Internet d'ingérence dans le droit à la liberté d'expression, et précise explicitement les critères de l'article 10 alinéa 2. En ce qui concerne l'emploi de filtres Internet par des entités privées, la recommandation et le rapport qui la sous-tend mentionnent les obligations positives incombant aux Etats membres au titre de garants du pluralisme, et en vue de préserver et de renforcer les mesures juridiques ainsi que pratiques visant à empêcher la censure privée.

Dans la troisième série de lignes directrices, la recommandation aborde spécifiquement les exigences légales s'appliquant aux filtres Internet à la lumière des dispositions de l'article 10 de la CEDH. Elle appelle les Etats membres à s'abstenir de filtrer le contenu de l'Internet sur les réseaux de communication électroniques gérés par des entités publiques pour des raisons autres que celles exposées à l'article 10, paragraphe 2. De même, elle déclare que les Etats membres ne sont autorisés à mettre en œuvre des mesures générales de blocage ou de filtrage sur tout le territoire, que si les conditions énoncées dans la CEDH, et notamment à l'article 10, paragraphe 2, sont remplies. Selon les lignes directrices de la recommandation, cela signifie que le filtrage doit concerner « un contenu spécifique et clairement identifiable », qu'« une autorité nationale compétente a pris une décision au sujet de l'illégalité de ce contenu » et que « la décision peut être réétudiée par un tribunal ou entité de régulation indépendant et impartial, en accord avec les dispositions de l'article 6 de la Convention européenne des Droits de l'Homme³⁷. » Les lignes directrices stipulent en outre que les Etats membres doivent veiller à ce que les filtres soient évalués avant et pendant leur mise en œuvre, afin de s'assurer que leurs effets sont justifiés au vu de possibles blocages excessifs des contenus.

Parmi les lignes directrices, plusieurs – dans la première et la troisième partie – concernent explicitement la gouvernance de l'inclusion et de l'exclusion de certains fournisseurs d'informations par les filtres Internet, ainsi que les droits des utilisateurs finaux et des fournisseurs d'informations dans ce domaine. Tout d'abord, les lignes directrices disposent que les utilisateurs finaux doivent être en mesure, le cas échéant, de contrôler le niveau de filtrage. Elles stipulent en outre que les utilisateurs finaux doivent avoir la possibilité de contester le blocage ou le filtrage du contenu, et de demander des explications ainsi que la mise en place de solutions. En ce qui concerne la capacité de l'utilisateur final, le cas échéant, à activer et à désactiver les filtres, et à recevoir une aide pour moduler le niveau du filtre en fonctionnement, les lignes directrices appellent les Etats membres à garantir, en coopération avec le secteur privé et la société civile, l'existence d'un certain nombre de conditions plus détaillées. Le texte affirme que les utilisateurs finaux doivent recevoir des conseils sur le contournement manuel d'un filtre actif – c'est-à-dire, notamment, sur l'instance à contacter quand le blocage d'un contenu semble injustifié, et sur les motifs qui peuvent autoriser le contournement d'un filtre pour un type spécifique de contenu ou localisateur universel de ressources (URL). En outre, la recommandation stipule que les contenus filtrés par mégarde ou par erreur doivent devenir accessibles sans difficulté indue et dans un délai raisonnable. Concernant l'utilisation et la mise en œuvre des filtres Internet par le secteur public,

37) Article 6, Convention européenne des Droits de l'Homme, 1^{er} paragraphe :

« Toute personne a droit à ce que sa cause soit entendue équitablement, publiquement et dans un délai raisonnable, par un tribunal indépendant et impartial, établi par la loi (...). Le jugement doit être rendu publiquement, mais l'accès de la salle d'audience peut être interdit à la presse et au public pendant la totalité ou une partie du procès dans l'intérêt de la moralité, de l'ordre public ou de la sécurité nationale dans une société démocratique, lorsque les intérêts des mineurs ou la protection de la vie privée des parties au procès l'exigent, ou dans la mesure jugée strictement nécessaire par le tribunal, lorsque dans des circonstances spéciales la publicité serait de nature à porter atteinte aux intérêts de la justice. »

les Etats membres doivent éviter le blocage général des contenus choquants ou préjudiciables pour les utilisateurs ne faisant pas partie du groupe qu'un filtre vise à protéger, ou pour les utilisateurs pouvant attester de l'intérêt ou de la nécessité légitime d'y accéder dans des circonstances exceptionnelles, notamment à des fins de recherche. Les Etats membres doivent de plus prévoir des voies de recours et des solutions effectives et facilement accessibles (dont la suspension des filtres), dans les cas où les usagers et/ou les auteurs de contenus dénoncent le blocage abusif d'un contenu.

2.4. Pertinence pour la gouvernance des outils de recherche

La recommandation du Conseil de l'Europe sur le filtrage Internet et le rapport d'experts qui la sous-tend offrent un point de départ intéressant dans l'analyse de l'incidence de la liberté d'expression sur la gouvernance de l'inclusion et de l'exclusion de certains fournisseurs d'informations par les moteurs de recherche ; ils abordent en effet les questions liées à la gamme de fournisseurs d'informations accessibles via un service de recherche donné.

Si l'on en croit la recommandation, les conséquences de la liberté d'expression seraient différentes concernant, d'une part, les obligations juridiques incombant aux moteurs de recherche pour limiter l'accès à certains fournisseurs d'informations et les moteurs de recherche fournis par le secteur privé, et, d'autre part, les limites posées volontairement à cet accès par les moteurs de recherche du secteur privé. Une obligation juridique imposée aux moteurs de recherche de filtrer certains fournisseurs d'informations et de les exclure de l'index de recherche doit respecter les conditions fixées par l'article 10, paragraphe 2. Si l'on en croit les lignes directrices, cela signifie que la suppression de certains fournisseurs d'informations de l'index du moteur de recherche doit porter sur des contenus spécifiques et clairement identifiables. Imposer des obligations de filtrage générales et préventives aux moteurs de recherche en ce qui concerne les contenus illégaux ou non autorisés pourrait donc être problématique de ce point de vue³⁸. Autre exigence intéressante, le fait que, dans les cas de filtrage obligatoire, une instance nationale compétente doive juger de l'illégalité de l'information, et qu'il faille offrir une possibilité de voir cette décision réétudiée par un tribunal ou une entité de régulation, compétent et impartial.

Les lignes directrices sur les filtres Internet visent à garantir un certain niveau de contrôle à l'utilisateur, ainsi que la transparence dans l'application des filtres Internet ; il est intéressant d'étudier comment elles s'appliqueraient dans le contexte des moteurs de recherche. Un premier problème apparaît ici : un moteur de recherche ne peut pas être activé ou désactivé à la façon d'un filtre Internet. Dans le cas des moteurs de recherche, l'utilisateur final d'Internet dispose d'autres possibilités d'action pour rétablir l'équilibre : il peut par exemple décider d'utiliser un autre moteur de recherche – un paramètre qui n'est différent que si le choix entre les moteurs de recherche est restreint, ce qui n'est normalement pas le cas. L'utilisateur final peut également accéder directement à l'information, mais cela suppose qu'il sache où se trouve l'information sur Internet. Dans bien des cas, si l'utilisateur a recours à un moteur de recherche, c'est parce qu'il ignore où elle se trouve. Les filtres complémentaires des résultats proposés par les moteurs de recherche (tels que les filtres d'images évoqués plus haut) offrent déjà aux utilisateurs un contrôle sur le filtrage dans les « paramètres avancés ». Ces possibilités pourraient s'accroître encore au fil des évolutions technologiques à venir.

Concernant les voies de recours et les solutions effectives et facilement accessibles, dans le cas d'un fournisseur d'informations qui ne serait pas inclus dans l'index d'un moteur de recherche, différents points valent d'être mentionnés. D'abord, les moteurs de recherche s'appuient en partie sur les suggestions de tierces parties, notamment des fournisseurs d'informations, pour choisir les sources à inclure dans leur index. Tous les grands moteurs de recherche du Web disposent de formulaires et d'outils en ligne pour soumettre des sites Web aux « crawlers ». Du fait des protocoles d'exclusion tels que *robots.txt*, on notera que le contrôle exercé par les fournisseurs d'informations sur leur indexation implique un contrôle moindre, pour les utilisateurs finaux, quant à la composition de l'index. En respectant des instructions telles que *robots.txt*, les grands moteurs de recherche privilégient les intérêts des fournisseurs d'information en leur permettant d'exercer un contrôle sur l'accessibilité de leurs contenus via un service de recherche, par rapport aux intérêts des utilisateurs finaux qui n'ont

38) Ceci semble conforme à l'article 15 de la Directive sur le commerce électronique (2000/31/CE). Il faut toutefois remarquer que l'article 15 ne s'applique pas aux moteurs de recherche, comme le montre l'article 21 de cette directive. Voir *infra*, 3^e partie.

pas accès à ce type de contrôle. Les moteurs de recherche généraux en sont arrivés à respecter ces protocoles pour diverses raisons, mais une présentation systématique de ces motifs dépasserait le champ du présent article. L'une des raisons principales est que les moteurs de recherche sur le Web ne se limitent pas aux métadonnées, mais stockent de grandes parties, voire la totalité des contenus des sites Web sur leurs serveurs. En respectant les protocoles d'exclusion, les moteurs de recherche échappent à certains – mais certainement pas à tous – les conflits potentiels avec les éditeurs en ligne³⁹. Dans son avis relatif aux moteurs de recherche, le Groupe de travail « Article 29 » recommande que les moteurs de recherche respectent ces protocoles, car ils pourraient être utilisés pour empêcher l'indexation des données à caractère personnel⁴⁰.

La situation peut être différente lorsqu'un moteur de recherche a retiré délibérément un site donné de son index. Ce retrait peut être dû au fait qu'un site Internet a recours à des techniques illégitimes pour influencer son classement dans les résultats. Le moteur de recherche peut également juger qu'il est de sa responsabilité de retirer certaines références de son index⁴¹. Autre possibilité, il peut être contraint par la loi ou par un juge de le faire dans le cas où une recherche produit des résultats illicites – par exemple parce qu'ils contiennent ou présentent un lien vers des informations contraires à la loi⁴². Ceci nous conduit à l'intéressante question des circonstances dans lesquelles les moteurs de recherche se retrouvent soumis à l'obligation légale de retirer certaines références de leur index. La question légale de la responsabilité des moteurs de recherche concernant les références à des contenus de tierces parties est fortement liée à celle de l'inclusion et de l'exclusion des fournisseurs d'informations dans un service de recherche.

3. Responsabilité des moteurs de recherche en matière de références illicites

Les moteurs de recherche peuvent notamment être amenés à retirer de leur index des références à des fournisseurs d'informations parce qu'il serait contraire à la loi de présenter ces références à leurs utilisateurs (par exemple parce qu'elles contiennent ou donnent accès à des contenus diffamatoires, ou constituent un lien vers des versions non autorisées de documents protégés par le droit d'auteur). Sans aucun doute, cette responsabilité des moteurs de recherche en matière de références illicites peut avoir un impact substantiel sur la composition de l'index des moteurs de recherche et, par là même, sur la libre circulation de l'information facilitée par les moteurs de recherche.

La question de la responsabilité des moteurs de recherche pour des contenus de tierces parties est en fait celle de la responsabilité des prestataires intermédiaires sur Internet⁴³. Dans l'Union euro-

39) L'un des conflits les plus notables entre éditeurs et moteurs de recherche est l'affaire belge *Copiepresse et al. v. Google et al.* – Tribunal de Première instance, 13 février 2007, *Computer Law Review International*, 2007-2, p.48-50, résumé et commentaire par Croux/Sasserath. L'affaire a fait l'objet d'un appel. Voir *Google loses copyright case launched by Belgian newspapers*, *International Herald Tribune*, 13 février 2007. Disponible sur : <http://www.iht.com/articles/ap/2007/02/13/business/EU-FIN-Belgium-Google-vs-Newspapers.php>, consulté le 1^{er} juillet 2008.

40) Voir *supra*, note 10, p.14 (« Il est essentiel que les fournisseurs de moteurs de recherche respectent le choix des éditeurs de sites Internet de ne pas participer à leurs services »).

41) En Allemagne, les grands moteurs de recherche ont conclu un accord d'auto-régulation afin de retirer certaines références de leur index « pour accroître la protection du consommateur, et pour protéger les enfants et la jeunesse contre des contenus inadaptés lorsqu'ils utilisent les moteurs de recherche ». Voir le Sous-code de conduite pour les fournisseurs de moteurs de recherche de l'organisation d'autorégulation volontaire des fournisseurs de services multimédia (« *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter* – FSM ») (VK-S), 21 avril 2004. Disponible sur : http://www.fsm.de/en/Subcode_of_Conduct_for_Search_Engine_Providers. Consulté le 1^{er} juillet 2008. Pour une présentation de cet accord, voir Wolfgang Schulz et Thorsten Held, *Der Index auf dem Index? Selbstzensur und Zensur bei Suchmaschinen*, dans Marcel Machill & Markus Beiler (ed.), *Die Macht der Suchmaschinen / The Power of Search Engines*, Köln, Halem, 2007, p. 71-87.

42) Pour un recueil de jurisprudence allemande et internationale concernant la responsabilité des moteurs de recherche en matière de références illicites, voir le travail de Stephan Ott chez Links & Law. Disponible sur : <http://www.linksandlaw.com/legalresources-publicationsenglish.htm> Consulté le 1^{er} juillet 2008.

43) Pour un article sur la responsabilité des prestataires intermédiaires en ligne, voir K.J. Koelman, *Liability for on-line intermediaries*, Imprimatur, IViR, Amsterdam, 1997 ; Rosa Julia-Barceló : *On-Line Intermediary Liability Issues: Comparing EU and US Legal Frameworks*, *European Intellectual Property Review*, 2000, Vol. 22 (3), 2000, p. 106-119 ; K.J. Koelman, *Online Intermediary Liability*, dans P.B. Hugenholtz (ed.), *Copyright and Electronic Commerce*, Information Law Series-8, Kluwer, 2000, p. 7-58.

péenne, cette responsabilité en ce qui concerne les contenus fournis par des tiers est abordée spécifiquement dans la Directive sur le commerce électronique⁴⁴. Cette dernière comprend, dans ses articles 12-15, des dispositions spéciales concernant les responsabilités des services de la société de l'information qui jouent le rôle de prestataires intermédiaires en ligne. La directive définit trois catégories de services intermédiaires, à savoir le « *simple transport* » (« *mere conduit* », article 12), le « *caching* » (article 13), et l'« hébergement » (article 14)⁴⁵. Pour chacune de ces catégories, elle prévoit une exonération de responsabilité conditionnelle « horizontale »⁴⁶, souvent appelée « *safe harbour* » en anglais. Ainsi, dans le cas des fournisseurs de services d'hébergement, la condition liée à cette exonération est que « (a) le prestataire n'ait pas effectivement connaissance de l'activité ou de l'information illicites et, en ce qui concerne une demande en dommages et intérêts, n'ait pas connaissance de faits ou de circonstances selon lesquels l'activité ou l'information illicite est apparente ; ou (b) le prestataire, dès le moment où il a de telles connaissances, agisse promptement pour retirer les informations ou rendre l'accès à celles-ci impossible ». L'exonération de responsabilité laisse explicitement ouverte la possibilité, pour un tribunal ou une autorité administrative, d'exiger du fournisseur qu'il empêche une action contraire à la loi ou y mette fin. Ces exonérations n'ont aucun effet sur le caractère licite du traitement des informations par les fournisseurs ou par tout autre type de services intermédiaires. Ce caractère licite doit être déterminé au regard des lois applicables de l'Etat membre. Pour finir, la directive proscribit l'imposition d'obligations générales aux prestataires de ces services visant à les faire surveiller les informations qu'ils transmettent ou stockent, ou rechercher des faits ou des circonstances révélant des activités illicites (article 15).

La directive se concentre sur les obligations juridiques des services de la société de l'information, ce qui inclut les moteurs de recherche⁴⁷. Toutefois, l'article 21, relatif à l'évaluation de la directive par la Commission européenne, montre que les moteurs de recherche (ou services offrant des outils de localisation) et les liens hypertextes ne sont pas concernés par le régime de responsabilité des prestataires intermédiaires présenté dans les articles 12-15 de la directive. L'article 21, paragraphe 2, dispose que « [le] rapport, en examinant la nécessité d'adapter la présente directive, analyse en particulier la nécessité de présenter des propositions relatives à la responsabilité des fournisseurs de liens d'hypertexte et de services de moteur de recherche [...] ». L'exclusion des moteurs de recherche de la réglementation sur la responsabilité des prestataires intermédiaires au niveau de l'UE a conduit à un traitement disparate des moteurs de recherche par les instances légiférantes et l'appareil judiciaire des différents Etats membres. Puisque la Directive sur le commerce électronique ne prévoit pas la pleine harmonisation des exonérations de responsabilité accordées aux prestataires intermédiaires, les Etats membres peuvent encore réglementer la responsabilité des fournisseurs de moteurs de recherche et des autres prestataires intermédiaires qui n'entrent pas dans le champ de la directive. Selon le premier rapport de la Commission européenne relatif à l'application de la Directive sur le commerce électronique, quelques Etats membres ont fait usage de cette possibilité, notamment l'Espagne, le Portugal, l'Autriche et le Liechtenstein (membre de l'EEE)⁴⁸. Les nouveaux Etats membres de l'UE que sont la Hongrie et la Pologne étendent également aux moteurs de recherche la limite de responsabilité dont bénéficient les fournisseurs de services d'hébergement. Les « *safe harbours* » accordés aux fournisseurs de moteurs de recherche qui existent en Espagne, au Portugal, en Hongrie et en Pologne ressemblent à celui qui existe pour les hébergeurs (article 14 de la directive). Ceux qui concernent les fournisseurs de moteurs de recherche en Autriche et au Liechtenstein rappellent celui qui est accordé pour le « *simple transport* » par l'article 12 de la directive.

44) Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (Directive sur le commerce électronique) JO L 178, 17 juillet 2000, p.1.

45) D'une manière générale, le « *simple transport* » concerne la transmission de contenus provenant d'un tiers, le « *caching* » décrit le stockage temporaire de contenus issus de tiers pour des raisons techniques, et l'« hébergement » fait référence au stockage en général de contenus fournis par un tiers sur des serveurs.

46) Cela signifie que les limites de responsabilité s'appliquent à la fois à la responsabilité civile et pénale, et couvrent la responsabilité en matière de droit d'auteur, de marque commerciale, de déclaration diffamatoire, de discours de haine, etc.

47) Comme on peut le lire dans le considérant 18 de la directive : « [...] Les services de la société de l'information ne se limitent pas exclusivement aux services donnant lieu à la conclusion de contrats en ligne, mais, dans la mesure où ils représentent une activité économique, ils s'étendent à des services qui ne sont pas rémunérés par ceux qui les reçoivent, tels que les services qui fournissent des informations en ligne ou des communications commerciales, ou ceux qui fournissent des outils permettant la recherche, l'accès et la récupération des données. », *ibid*, p. 6.

48) Voir le « Premier rapport sur l'application de la Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur », 21 novembre 2003, COM (2003)702, p. 14.

Résultat de cette absence d'exonération de responsabilité à l'échelle européenne pour les fournisseurs de moteurs de recherche, et de ce manque d'harmonisation, on trouve en Europe un patchwork de lois régissant la responsabilité des fournisseurs d'outils de recherche en matière de références illicites. La question est de savoir si cette situation, qui implique un certain degré d'incertitude juridique, est favorable du point de vue du rôle facilitateur que jouent les services de moteurs de recherche en matière de libre circulation de l'information. Il va sans dire que l'imposition aux moteurs de recherche d'une obligation renforcée de diligence pourrait avoir un effet réfrigérant sur la circulation des informations licites, et pourrait aboutir à l'absence totale de moteurs de recherche efficaces⁴⁹. Dans l'idéal, il faudrait que la loi prévoie des incitations adaptées pour que les moteurs de recherche retirent les résultats illicites, sans pour autant figer la circulation d'informations licites facilitée par les moteurs de recherche.

On connaît quelques exemples de juridictions nationales prenant en compte l'importance des moteurs de recherche pour la mise en œuvre concrète de la liberté d'expression et d'information dans ce domaine. Invoquant à la fois l'utilité sociale des moteurs de recherche et le droit fondamental à la liberté d'expression et d'information, le jugement Paperboy de la Cour fédérale de justice allemande (*Bundesgerichtshof*) a reconnu aux moteurs de recherche utilisant des robots indexeurs le droit d'avoir recours aux technologies hypertexte⁵⁰. Dans une décision plus récente, la cour d'appel de Hambourg (*Oberlandesgericht*) prolonge ce jugement et conclut que la norme de responsabilité régissant les publications est trop stricte dans le cas des références publiées par les moteurs de recherche ; elle demande une exception à leur égard⁵¹. La cour d'appel conclut que la responsabilité des moteurs de recherche, concernant les contenus potentiellement illicites pouvant figurer parmi leurs références, est le fruit de la reproduction automatique d'informations illicites provenant de milliards de sites sur le Web, et doit donc être limitée en raison de l'importance des moteurs de recherche pour la liberté d'expression et d'information.

On ignore à ce jour si une forme de « *safe harbour* » sera adoptée au niveau européen en faveur des fournisseurs de moteurs de recherche. L'article 21 de la directive enjoint à la Commission européenne de mener tous les deux ans un rapport sur l'application de la directive, rapport qui doit comprendre une analyse du besoin de propositions en matière de responsabilité des fournisseurs de liens hypertextes et d'outils de localisation de l'information. Dans le premier rapport, la commission est parvenue à la conclusion qu'il n'y avait pas de raison d'amender les règles existantes quant aux responsabilités intermédiaires en matière de moteurs de recherche. La commission a bien pris note des choix législatifs divergents, et écrit ce qui suit au sujet des évolutions juridiques en matière de responsabilité des moteurs de recherche :

« Il est encourageant de constater que la jurisprudence récente des États membres reconnaît l'importance des liens hypertexte et des moteurs de recherche pour le fonctionnement d'Internet. En général, cette jurisprudence semble aller dans le même sens que l'objectif du marché intérieur d'assurer la fourniture de services intermédiaires fondamentaux facilitant le développement de l'Internet et du commerce électronique. Par conséquent cette jurisprudence ne devrait donc pas soulever de préoccupations quant au bon fonctionnement du marché intérieur⁵². »

La commission a par suite inclus cette question dans le programme du réexamen suivant qui, selon l'article 21 de la directive, aurait dû être finalisé en 2005. Dans le deuxième rapport, une attention particulière sera portée à la responsabilité des prestataires intermédiaires, et notamment à celle des moteurs de recherches en matière de contenus provenant de tiers. Dans son réexamen, la Commission européenne devra également examiner cette question du point de vue de la liberté d'expression. Si la logique de base de la directive est le renforcement du Marché intérieur pour le commerce électronique et les services afférents, la directive admet que « dans bien des cas, la libre circulation des services de

49) A ce sujet, voir Koelman dans P.B. Hugenholtz (ed.), *op. cit.*, p. 40-44

50) *Bundesgerichtshof* (Cour fédérale de justice allemande), 17 juillet 2003, I ZR 259/00.

51) *Oberlandesgericht* de Hambourg, 20 février 2007 – AZ. : 7 U 126/06.

52) Premier rapport sur l'application de la Directive 2000/31/CE du Parlement européen et du Conseil du 8 juin 2000 relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur, 21 novembre 2003, COM(03)0702, p. 13. Disponible sur : http://ec.europa.eu/internal_market/e-commerce/directive_fr.htm

la société de l'information peut refléter spécifiquement, dans la législation communautaire, un principe plus général, à savoir la liberté d'expression, consacrée par l'article 10, paragraphe 1, de la Convention de sauvegarde des droits de l'homme et des libertés fondamentales, qui a été ratifiée par tous les Etats membres [...]»⁵³. »

4. Conclusion

Les moteurs de recherche ont acquis une place centrale dans l'environnement informationnel en ligne et jouent un rôle crucial de facilitation des communications publiques sur Internet. Ils assurent une médiation entre les intérêts des fournisseurs d'informations et des utilisateurs finaux liés à la libre circulation des informations en ligne, et ont un impact substantiel sur l'accès à l'information – y compris aux contenus audiovisuels. En raison de leur rôle d'auxiliaire dans la libre circulation de l'information, et, plus spécifiquement, dans la capacité des fournisseurs d'informations à atteindre leur public et dans les intérêts des utilisateurs finaux à rechercher le contenu de leur choix, le présent article s'est concentré sur l'incidence des droits fondamentaux à la liberté d'expression et d'information, tels qu'ils sont consacrés par l'article 10 de la Convention européenne sur les droits de l'homme, sur la gouvernance des moteurs de recherche. Nous avons étudié certaines questions générales liées à cette incidence – notamment concernant la gouvernance de l'inclusion et de l'exclusion des fournisseurs d'informations dans l'index d'un moteur de recherche. Nous avons montré que la recommandation récente du Conseil de l'Europe sur les filtres Internet constituait un point de départ intéressant pour étudier ces questions. Parce qu'il s'agit d'un sujet connexe, la responsabilité des moteurs de recherche en matière de contenus fournis par des tiers au niveau de l'UE a également été abordée. Le régime de responsabilité en vigueur et la faible certitude juridique peuvent être considérés comme étant sous-optimaux du point de vue de la liberté d'expression et d'information. Globalement, les défis juridiques et réglementaires concernant les moteurs de recherche peuvent être qualifiés de substantiels. Les moteurs de recherche sont des outils complexes, aux facettes multiples, et seul le temps permettra de répondre à certaines des questions soulevées dans cet article.

53) Directive 2000/31/CE, considérant 9.

À la recherche des moteurs de recherche audiovisuels dans le cadre réglementaire de l'UE

*Peggy Valcke**

*Centre interdisciplinaire pour le droit et les technologies de l'information
(Interdisciplinary Centre for Law & ICT) de l'Université catholique de Louvain
Université catholique de Bruxelles.*

Introduction

Dans un rapport publié dans *IRIS plus* 2006-02, Nico van Eijk affirme que « le moteur de recherche n'a pas sa place dans le droit¹. » Il impute le peu d'attention accordée jusqu'à présent par les législateurs² aux moteurs de recherche au fait qu'ils se trouvent dans la zone d'ombre située entre les communications électroniques (transmission) et la radiodiffusion (contenu). Depuis, la Directive « Télévision sans frontières »³ (voir plus bas 1.3.) pierre angulaire de la réglementation communautaire en matière de radiodiffusion, et les directives sur les communications électroniques⁴ (voir plus bas 1.2.) ont fait, ou font encore, l'objet d'une révision approfondie. Cela a-t-il permis (ou permet-il aujourd'hui) de changer le traitement juridique réservé aux moteurs de recherche, ou plus généralement, les outils de navigation ? L'objet du présent rapport est d'analyser le cadre réglementaire de l'Union européenne (UE) et d'examiner dans quelle mesure les différents dispositifs sont applicables, le cas échéant, aux moteurs de recherche audiovisuels et aux outils de navigation⁵. Dans la seconde partie, nous formulerons quelques propositions visant à « combler les lacunes ».

Avertissement

Je tiens à apporter, en préambule, les précisions suivantes : l'extrême complexité de cette analyse réside dans le fait qu'il existe toute une variété d'outils de recherche audiovisuels. Certains moteurs

* Peggy Valcke est maître de conférences en droit de la communication et des médias à l'Université catholique de Louvain et de l'Université catholique de Bruxelles et elle a été professeur associée à l'Université d'Europe centrale de Budapest en 2006-2007. Elle travaille comme chercheuse postdoctorale au Centre interdisciplinaire pour le droit et les technologies de l'information (Interdisciplinary Centre for Law & ICT) de l'Université catholique de Louvain (www.icri.be), qui est membre fondateur de l'Institut flamand des technologies de radiodiffusion (Flemish Institute for Broadband Technology - www.ibbt.be).

1) Nico Van Eijk, 2006 Moteurs de recherche : Cherche et tu trouveras ? Positionnement juridique des moteurs de recherche. *IRIS plus* 2006-2 ; Strasbourg : Observatoire européen de l'audiovisuel, disponible sur : http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf

2) Contrairement aux tribunaux, comme en attestent les nombreuses décisions de justice concernant les moteurs de recherche dans plusieurs domaines juridiques (propriété intellectuelle, pratiques commerciales déloyales, protection des données ...).

3) Directive 89/552/CEE du Conseil, du 3 octobre 1989, visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à l'exercice d'activités de radiodiffusion télévisuelle, [1989] JO L298/23, dans sa version amendée par la Directive 97/36/CE, [1997] JO L202/60.

4) En bref : Directive 2002/21/CE (Directive « cadre ») [2002] JO L108/33, Directive 2002/19/CE (Directive « accès ») [2002] JO L108/7, Directive 2002/20/CE (Directive « autorisation ») [2002] JO L108/21, Directive 2002/22/CE (Directive « service universel ») [2002] JO L108/51, Directive 2002/58/CE (Directive « vie privée et communications électroniques ») [2002] JO L201/37. Ces directives sont complétées par la Directive 2002/77/CE de la Commission (Directive « concurrence » [2002] JO L249/21) et par un certain nombre de réglementations, décisions et recommandations, (pour des informations plus détaillées, voir : http://ec.europa.eu/information_society/policy/ecomm/current/index_en.htm).

5) Dans cet article, les termes « moteurs de recherche audiovisuels » et « outils de navigation » sont synonymes.

recherchent du matériel audiovisuel sur le Web, dans le cadre d'un site de recherche plus général ou horizontal (englobant textes, images, fichiers stockés localement, etc., comme Exalead⁶) ou dans le cadre d'un service de *bookmarking* social (par ex. del.icio.us⁷). D'autres sont dédiés à la recherche de vidéos, en associant parfois la fonction de recherche à la capacité d'hébergement (gratuit) pour les utilisateurs qui souhaitent télécharger leurs propres vidéos (par ex. GoogleVideo⁸). Dans d'autres cas, la fonctionnalité première n'est plus la recherche, mais l'hébergement et la distribution, comme c'est le cas avec la prolifération des portails vidéo en ligne. Parmi ces derniers, certains sont axés sur les contenus générés par les utilisateurs (tels que YouTube, Dailymotion, GarageTV or BlipTV⁹) et d'autres sur les chaînes, émissions, films et clips vidéos professionnels (comme Joost¹⁰). D'autres portails, quant à eux, combinent ces deux types de contenus (comme Gotuit Media¹¹) et/ou offrent aux utilisateurs des outils de communication et d'évaluation pour partager leur expérience avec des amis tout en regardant les contenus (par ex. les logiciels de messagerie instantanée ou un espace blog)¹².

À l'autre extrémité de l'éventail, on trouve les guides électroniques de programmes conventionnels (édités – EPG) utilisés dans un environnement de type *walled-garden* (pré carré d'une gamme de services proposés) de la télévision numérique. À présent que les radiodiffuseurs établis sortent des plateformes de télévision classiques pour étendre leur champ d'action à Internet, on trouve de plus en plus de guides et de répertoires de programmes en ligne qui permettent aux utilisateurs de savoir ce qui est diffusé sur telle chaîne à tel moment, en soirée ou le lendemain (la BBC¹³ en est un exemple parmi tant d'autres). De nouveaux portails de divertissement (comme le site irlandais entertainment.ie¹⁴) et certains magazines de programmes télévisés proposent également des présentations similaires en ligne, qui reproduisent leurs versions imprimées sur Internet ; ces programmes en ligne sont souvent assortis de fonctionnalités supplémentaires et généralement gratuits, c'est-à-dire financés par la publicité et non par l'abonnement des utilisateurs, comme par exemple TV Guide¹⁵. Ces portails ont généralement pour pratique de présenter une grille des programmes diffusés à un moment précis sur les différentes chaînes, en offrant la possibilité aux utilisateurs de personnaliser leur grille en fonction de leurs préférences (par ex. en modifiant l'ordre des chaînes), de rechercher une émission particulière ou de consulter des informations supplémentaires sur les programmes. Certains proposent même un service de programmation en ligne, c'est-à-dire un outil permettant d'enregistrer des programmes sur son magnétoscope numérique par un simple clic dans la grille de programmes (comme TVGuide, par ex., qui s'est associé pour cela avec TiVo[®] DVR). Contrairement aux guides électroniques de programmes sur les plateformes télévisées classiques, ces portails n'offrent pas¹⁶, du moins pour l'instant, la possibilité d'accéder à la chaîne ou au programme directement à partir de la grille en ligne. Au vu de la convergence croissante entre Internet et la télévision, ce n'est plus qu'une question de temps avant que les guides de programmes en ligne ne soient en mesure de proposer l'accès et le zapping ; dans le même temps, les guides électroniques de programmes « conventionnels » intégreront des fonctionnalités comparables qui permettront d'accéder aux contenus générés par les utilisateurs et disponibles sur Internet et d'y faire des recherches.

Autrement dit, il existe un continuum presque infini de services et de produits de recherche pour aider l'utilisateur à s'orienter dans le « foisonnement audiovisuel ». Mais cette fonction commune de recherche ne signifie pas automatiquement que tous ces services soient assujettis au même régime juridique (ce n'est pas parce qu'un véhicule a des roues qu'il faut un permis pour le conduire). La situation est fondamentalement différente selon que la fonction de recherche fait partie d'une offre de contenus plus large (bien que, dans ce cas, le prestataire risque d'être largement « épargné » si l'offre

6) <http://www.exalead.com>

7) <http://del.icio.us>

8) <http://video.google.com/>

9) <http://youtube.com/> , <http://www.dailymotion.com/> , <http://www.garagetv.be/> , <http://blip.tv>

10) <http://www.joost.com>

11) <http://www.gotuit.com>

12) Une analyse intéressante de ces nouveaux portails sous l'angle de la régulation des contenus nous est proposée par Monica Ariño, 2007 : *Content Regulation and New Media: A Case Study of Online Video Portals. Communications and Strategies* (66): 115 – 135.

13) <http://www.bbc.co.uk/cgi-perl/whatson/search/grid.cgi?MEDIUM=tv>

14) <http://entertainment.ie/tv/whats-on-now.asp>

15) <http://www.tvguide.com/Listings/default.aspx> (un exemple similaire tiré du contexte flamand : <http://www.tvguids.be/>)

16) Du moins ceux que nous avons examinés.

s'apparente plutôt à l'équivalent électronique de produits imprimés ou de textes), qu'elle relève d'une offre de contenus audiovisuels plus vaste (auquel cas le service de recherche peut être soumis aux mêmes règles que le service de médias audiovisuels qu'il accompagne, sauf si le service de contenus ne remplit pas tous les critères de la définition juridique d'un service de média audiovisuel – voir ci-dessous), ou qu'elle soit intégrée dans des services de transmission (comme les EPG sur les plateformes classiques de télévision qui sont proposées par les fournisseurs de services d'accès conditionnel – voir ci-dessous).

Cet article n'a pas pour vocation d'offrir une analyse juridique complète de toutes les formes possibles d'outils de recherche audiovisuels. L'objectif est de mettre en évidence les dispositions pertinentes dans le cadre réglementaire de l'UE (c'est-à-dire les dispositions faisant référence aux moteurs de recherche ou aux guides électroniques de programmes), d'analyser les concepts utilisés et d'étudier la portée des règles qui en découlent.

1. Les outils de navigation dans la réglementation communautaire du secteur des infocommunications

1.1. À la recherche des moteurs de recherche...

Les moteurs de recherche audiovisuels sont régis par une série de dispositions juridiques générales de l'UE, notamment les règles en matière de concurrence des articles 81 et 82 du Traité CE, les règles relatives aux pratiques commerciales déloyales de la Directive 2005/29/CE¹⁷, les règles sur la publicité prévues, par exemple, par les Directives 2006/114/CE (publicité trompeuse et comparative)¹⁸, 2003/33/CE (publicité sur le tabac)¹⁹ et 2001/83/CE (produits médicaux)²⁰. Toutefois, ces outils ne seront pas étudiés dans le cadre de cet article.

L'objet du présent rapport est de se pencher sur les règles adoptées spécifiquement pour les secteurs de la radiodiffusion et des services audiovisuels, des communications et de l'information, regroupés sous le terme de « secteur des infocommunications ». La réglementation sectorielle de l'UE en matière d'infocommunication est basée sur la distinction entre transmission et contenu. Les directives sur les communications électroniques adoptées en 2002 couvrent tous les réseaux et services de transmission (résultat de la convergence), mais non les services de contenus fournis par le biais de ces réseaux (tels que les contenus radiodiffusés, les services financiers et certains services propres à la société de l'information²¹). Les services liés aux contenus, qui consistent à mettre en forme et fournir des contenus audiovisuels ou des informations, sont régis par la Directive « Télévision sans frontières »²² (récemment modifiée et rebaptisée Directive sur les services de médias audiovisuels, voir ci-dessous) et par la Directive sur le commerce électronique²³. Cependant, les imbrications complexes entre transmission et contenus ne permettent pas de réglementer l'une et l'autre de façon totalement dissociée, ni d'exclure des mesures hybrides ou « intermédiaires ».

Le cadre des communications électroniques, par exemple, comporte des règles qui s'appliquent aux réseaux de transmission ou aux ressources associées pour des raisons liées au contenu (telles que la

17) Directive 2005/29/CE du Parlement européen et du Conseil du 11 mai 2005 relative aux pratiques commerciales déloyales des entreprises vis-à-vis des consommateurs dans le marché intérieur et modifiant la Directive 84/450/CEE du Conseil et les Directives 97/7/CE, 98/27/CE et 2002/65/CE du Parlement européen et du Conseil et le règlement (CE) n° 2006/2004 du Parlement européen et du Conseil (« Directive sur les pratiques commerciales déloyales ») [2005] JO L 149/22.

18) Directive 2006/114/CE du Parlement européen et du Conseil du 12 décembre 2006 en matière de publicité trompeuse et de publicité comparative [2006] JO L 376/21.

19) Directive 2003/33/CE du Parlement européen et du Conseil du 26 mai 2003 concernant le rapprochement des dispositions législatives, réglementaires et administratives des États membres en matière de publicité et de parrainage en faveur des produits du tabac, [2003] JO L 152/16.

20) Directive 2001/83/CE du Parlement européen et du Conseil du 6 novembre 2001 instituant un code communautaire relatif aux médicaments à usage humain, [2001] JO L 311/67, dans sa version amendée par la Directive 2004/27/CE du 31 mars 2004, [2004] JO L 136/34.

21) Cf. considérant 5 de la Directive « cadre ».

22) Cf. ci-dessus note 3 et ci-après note 39.

23) Cf. ci-après note 47.

diversité culturelle) ; rappelons, à cet égard, les dispositions sur l'obligation de distribution (*must-carry*) visées à l'article 31 de la Directive sur le service universel ou les règles concernant les dispositifs de goulots d'étranglement pour la télévision numérique (voir ci-dessous).

1.2. ...dans les directives sur les communications électroniques

Le cadre réglementaire de 2002

Comme nous l'avons mentionné précédemment, les directives de 2002 sur les communications électroniques s'appliquent aux réseaux et services de transmission ainsi qu'aux « ressources associées ». Étant donné que les outils de recherche ne constituent manifestement pas un réseau en tant que tels, nous pouvons limiter notre analyse aux concepts des services de communications électroniques et aux ressources associées.

Les premiers sont définis par l'article 2 c) de la Directive « cadre » comme « les services fournis normalement contre rémunération qui consistent entièrement ou principalement en la transmission de signaux sur des réseaux de communications électroniques, y compris les services de télécommunications et les services de transmission sur les réseaux utilisés pour la radiodiffusion ». La définition exclut de façon formelle « les services consistant à fournir des contenus à l'aide de réseaux et de services de communications électroniques ou à exercer une responsabilité éditoriale sur ces contenus », de même que « les services de la société de l'information tels que définis à l'article 1^{er} de la Directive 98/34/CE qui ne consistent pas entièrement ou principalement en la transmission de signaux sur des réseaux de communications électroniques ». Ces exceptions semblent rédigées à l'égard des services radiodiffusés et audiovisuels à la demande, en excluant ceux qui mettent en forme ce type de service, de même que ceux qui fournissent (ou distribuent) ce type de contenu au public. Ni la mise en forme, ni la fourniture ou la distribution des contenus via Internet ou des plateformes plus classiques, comme la télévision par câble, ne sont considérées comme des services de communications électroniques²⁴. Les outils de recherche font office de services de répertoire qui facilitent l'accès aux informations de tiers, néanmoins ils n'éditent pas eux-mêmes les informations ou les contenus et, d'autre part, on peut difficilement affirmer qu'ils procurent eux-mêmes les contenus. En même temps, il est indéniable que les services qu'ils fournissent ne consistent pas entièrement ou principalement en la transmission de signaux. Leur activité de base ne consiste pas simplement à transporter des signaux du point A au point B, mais d'aider l'utilisateur à trouver les informations qu'il cherche en référençant les informations de tiers, classées en fonction de leur pertinence sur la base des termes de la recherche et/ou des préférences antérieures de l'utilisateur. Ainsi, même si les moteurs de recherche comportent certains aspects des services de transmission, leur valeur ajoutée réside dans leurs fonctionnalités en tant qu'outils de marketing et de services liés aux contenus ce qui, par conséquent, les exclut du champ des services de communications électroniques au sens visé par l'article 2 c) de la Directive « cadre »²⁵.

La notion de ressources associées est définie à l'article 2 e) de la Directive « cadre » comme « les ressources associées à un réseau de communications électroniques et/ou à un service de communications électroniques, qui permettent et/ou soutiennent la fourniture de services via ce réseau et/ou ce service; elles comprennent les systèmes d'accès conditionnel et les guides électroniques de programmes. » Par conséquent, les EPG qui sont une forme d'outil de recherche audiovisuel, comme nous l'avons expliqué dans l'introduction, sont explicitement couverts par ces directives. Mais jusqu'où cette notion peut-

24) Voir également le considérant 10 de la Directive « cadre » (« La même entreprise, par exemple un prestataire de services Internet, peut proposer à la fois un service de communications électroniques, tel que l'accès à Internet, et des services non couverts par la présente directive, tels que la fourniture de contenus sur la toile ») et le considérant 20 de la Directive « autorisation » (Une même entreprise (par exemple, un câble-opérateur) pouvant offrir à la fois un service de communications électroniques, comme l'acheminement de signaux télévisés, et des services non couverts par la présente directive, comme la commercialisation d'une offre de contenus de radio ou de télédiffusion, des obligations supplémentaires peuvent lui être imposées à propos de son activité de fournisseur ou de distributeur de contenus, conformément à des dispositions autres que celles de la présente directive, sans préjudice de la liste de conditions figurant à l'annexe de cette dernière »).

25) Cf. Nico van Eijk (2006) sur les moteurs de recherche. Voir également l'avis récent du groupe de travail Article 29 sur les aspects de la protection des données liés aux moteurs de recherche, mentionné ci-après à la note 57 : « Par conséquent, les moteurs de recherche n'entrent pas dans le champ de la définition des services de communications électroniques » (ce qui, néanmoins, n'empêche pas l'exploitant d'un moteur de recherche d'être assujéti à la Directive sur les communications électroniques s'il propose des services complémentaires qui correspondent à la définition des services de communications électroniques, tels que, par exemple, un service de messagerie électronique d'accès public).

elle s'étendre ? Pourrait-elle, par exemple, englober les moteurs de recherche sur Internet ? Contrairement aux concepts des systèmes d'accès conditionnel (CAS) et aux interfaces de programme d'application (API)²⁶, les EPG ne sont pas définis plus en détail. Ce n'est qu'en examinant les règles correspondantes de la Directive « accès » qu'il apparaît clairement²⁷ que le concept d'EPG est également limité au cadre des services de transmissions radiophoniques et télévisées numériques et uniquement règlementé du point de vue des aspects techniques (transmission) :

- L'article 5, paragraphe 1 b) et l'article 6 de la Directive « accès » (en lien avec l'annexe I) prévoient un régime d'accès spécifique pour les ressources associées, et plus particulièrement pour les systèmes d'accès conditionnel (article 6), les interfaces de programme d'application et les guides électroniques de programmes *dans le domaine des services de transmissions radiophoniques et télévisées numériques*²⁸. De ce fait, il est peu probable que les moteurs de recherche sur Internet puissent être considérés comme des guides électroniques de programmes au sens où l'entendent les directives sur les communications électroniques.
- L'objectif de ces dispositions est de fournir des outils garantissant l'accès à ces ressources dans des conditions équitables, raisonnables et non discriminatoires (ERND). L'accès doit être compris au sens technique, ici, en référence exclusive aux aspects de transmission (accès à la technologie mise en œuvre ou aux ressources de transport uniquement). Nous pouvons déduire cela non seulement de la définition de l'« accès » proprement dit visée à l'article 2 a) de la Directive « accès »²⁹, mais aussi de l'article 6, alinéa 4, qui spécifie que « les conditions fixées au titre du présent article sont appliquées sans préjudice de la possibilité laissée aux États membres d'imposer des obligations en rapport avec la présentation des guides électroniques de programmes et des outils de présentation et de navigation similaires. » Par conséquent, les États membres restent libres de réglementer les aspects liés aux contenus des guides électroniques de programmes et des « outils de présentation et de navigation similaires » (tels que les moteurs de recherche sur Internet) puisqu'ils ne sont pas couverts par les directives sur les communications électroniques.

Les outils de recherche autres que les guides électroniques de programmes pourraient-ils être néanmoins inclus dans la catégorie plus large des « ressources associées » ? Éventuellement, puisque cette notion est formulée de façon large (sans être limitée au secteur de la radiodiffusion numérique) et qu'on peut affirmer que les outils de recherche « permettent et/ou soutiennent la fourniture de services via ce réseau et/ou ce service » (exactement comme le font les guides de programmes électroniques dans le domaine de la radiodiffusion numérique). A cet effet, il est important de noter que la notion de ressources associées a été intentionnellement étendue durant le processus législatif visant à l'adoption de la Directive « cadre » en 2002, de ressources nécessaires pour fournir des « services de communications électroniques » à des ressources permettant ou encourageant la fourniture de « services » (incluant la radiodiffusion ou les services de la société de l'information)³⁰.

26) Cf. respectivement articles 2 (f) et 2 (p) de la Directive « cadre ».

27) Pour les systèmes d'accès conditionnel et les interfaces de programmes d'application, cela apparaît clairement dans la définition elle-même.

28) L'article 5(1) et l'annexe 1, partie II de la Directive « accès » spécifient qu'en l'absence de puissance significative sur le marché, les autorités réglementaires nationales doivent être en mesure d'imposer aux opérateurs, dans la mesure de ce qui est nécessaire pour assurer l'accès des utilisateurs finals à des services de transmissions radiophoniques et télévisées numériques spécifiés par l'État membre, l'obligation de fournir l'accès à des API et des EPG dans des conditions équitables, raisonnables et non discriminatoires.

29) « accès » : la mise à la disposition d'une autre entreprise, dans des conditions bien définies et de manière exclusive ou non exclusive, de ressources et/ou de services en vue de la fourniture de services de communications électroniques. Cela couvre notamment: l'accès à des éléments de réseaux et à des ressources associées et éventuellement la connexion des équipements par des moyens fixes ou non (cela inclut en particulier l'accès à la boucle locale ainsi qu'aux ressources et services nécessaires à la fourniture de services via la boucle locale) ; l'accès à l'infrastructure physique, y compris les bâtiments, gaines et pylônes; l'accès aux systèmes logiciels pertinents, avec notamment les systèmes d'assistance à l'exploitation ; l'accès à la conversion du numéro d'appel ou à des systèmes offrant des fonctionnalités équivalentes ; l'accès aux réseaux fixes et mobiles, notamment pour l'itinérance ; l'accès aux systèmes d'accès conditionnel pour les services de télévision numérique ; l'accès aux services de réseaux virtuels. »

30) Voir par exemple le rapport Paasilinna (ITRE) sur la proposition de directive du Parlement européen et du Conseil relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques (COM(2000 393-C5-0428/2000 – 2000/0184(COD)), A5-0053/2001, PE 297.096, 7 février 2001, amendement 19, p. 17 et 18.

L'utilité d'une telle qualification reste toutefois limitée. Même si le régime général d'une position de puissance significative sur le marché peut, en principe, s'appliquer aux ressources associées, il est très improbable que soit déclenchée, à court ou moyen terme, l'application des obligations a priori aux fournisseurs de ces ressources ayant une position de nature dominante sur le marché. Tout d'abord, parce qu'un tel marché n'est pas compris dans les Recommandations de la Commission sur les marchés pertinents³¹, ce qui signifie que les autorités de régulation nationales qui voudraient l'analyser se heurteraient à une charge considérable de la preuve, dans la mesure où il leur faudrait démontrer que les conditions du « test des trois critères » sont réunies³². Ensuite, les recours pouvant être appliqués se réfèrent toujours à la notion « d'accès » au sens technique (comme nous l'avons précédemment expliqué) qui ne couvre que les aspects de la transmission et ne sont donc pas (bien) en mesure de traiter les problèmes liés aux contenus.

Propositions de réforme

Les directives sur les communications électroniques sont en cours de révision. La Commission européenne a publié ses propositions législatives le 13 novembre 2007 et le Parlement européen les a examinés en première lecture le 24 septembre 2008³³. Aucun changement fondamental n'a été introduit en ce qui concerne les « ressources associées » ou les fonctions de distribution, même si plusieurs suggestions ont été émises par certains consultants dans le cadre des études préliminaires. L'étude Cullen, en 2006, par exemple, a souligné la nécessité de clarifier la fonction de distribution au niveau du cadre réglementaire des communications électroniques (qui est en majeure partie dans une zone d'ombre)³⁴.

La notion de « ressources associées » restera identique en ce qui concerne les guides électroniques de programmes. Ces guides ne sont pas concernés par les modifications proposées par la Commission à l'article 2^e de la Directive « cadre » : les « ressources associées » désignent les ressources associées à un réseau de communications électroniques et/ou à un service de communications électroniques, qui permettent et/ou soutiennent la fourniture de services via ce réseau et/ou ce service **ou qui en ont le potentiel** ; elles comprennent **les systèmes de conversion des numéros d'appels ou des adresses**, les systèmes d'accès conditionnel et les guides électroniques de programmes, **ainsi que l'infrastructure physique, notamment les gaines, les pylônes, les boîtiers situés dans la rue et les bâtiments** »³⁵). Le Parlement européen a adopté un amendement rajoutant les ressources suivantes, qui n'ont aucun lien avec les guides électroniques de programme : « **entrées des bâtiments, le câblage des bâtiments, les tours et autres constructions de soutènement**, les gaines, **les conduites, les pylônes, les antennes, les trous de visite et les boîtiers et tous les autres éléments de réseau qui ne sont pas actifs.** » Les changements (mineurs) suggérés aux articles 5 et 6 de la Directive « accès » ne modifient pas le statut légal des guides électroniques de programme.

31) Recommandation de la Commission du 17 décembre 2007 concernant les marchés pertinents de produits et de services dans le secteur des communications électroniques susceptibles d'être soumis à une réglementation *ex ante* conformément à la Directive 2002/21/CE du Parlement européen et du Conseil relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques, [2007] JO L344/65.

32) Pour recenser les marchés susceptibles d'être soumis à une réglementation *ex ante*, il y a lieu d'appliquer les trois critères cumulatifs suivants : le premier critère est la présence de barrières élevées et non provisoires à l'entrée, le deuxième critère porte sur l'absence de structure dynamique du marché susceptible d'évoluer vers une situation de concurrence effective dans un délai adéquat et le troisième critère réside dans l'incapacité du droit de la concurrence à remédier à lui seul à la ou aux défaillance(s) concernée(s) du marché. Si un marché est inclus dans la liste de la Commission, on peut présumer qu'il répond à l'ensemble de ces critères ; dans le cas contraire, l'autorité de régulation nationale doit fournir la preuve correspondante.

33) Résolutions législatives du Parlement européen du 24 septembre 2008 sur la proposition de directive du Parlement européen et du Conseil modifiant les Directives 2002/21/CE relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques, 2002/19/CE relative à l'accès aux réseaux de communications électroniques et aux ressources associées ainsi qu'à leur interconnexion, et 2002/20/CE relative à l'autorisation des réseaux et services de communications électroniques, 6_TA(2008)0449 (COM(2007)0697 – C6-0427/2007 – 2007/0247(COD)) – A6-0321/2008 et sur la proposition de directive du Parlement européen et du Conseil modifiant la Directive 2002/22/CE concernant le service universel et les droits des utilisateurs au regard des réseaux et services de communications électroniques, la Directive 2002/58/CE concernant le traitement des données à caractère personnel et la protection de la vie privée dans le secteur des communications électroniques et le règlement (CE) n° 2006/2004 relatif à la coopération en matière de protection des consommateurs, P6_TA(2008)0452 (COM(2007)0698 – C6-0420/2007 – 2007/0248(COD)) – A6-0318/2008.
Disponible sur : <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//NONSGML+TA+20080924+SIT+DOC+WORD+V0//FR&language=FR>

34) Cullen International. 2006. *Study on the Regulation of Broadcasting Issues under the New Regulatory Framework*, p. 11-14.

35) Le texte en gras indique les modifications proposées.

La Commission, cependant, a proposé d'étendre la notion d'« accès » visée à l'article 2 e) de la Directive « accès » comme suit : « Le terme « accès » désigne la mise à la disposition d'une autre entreprise, dans des conditions bien définies et de manière exclusive ou non exclusive, de ressources et/ou de services en vue de la fourniture de services de communications électroniques, **de services propres à la société de l'information, ou de services de contenus radiodiffusés**. Cela couvre notamment : l'accès à des éléments de réseaux et à des ressources associées [...], l'accès à l'infrastructure physique, y compris les bâtiments, gaines et pylônes ; l'accès aux systèmes logiciels pertinents, avec notamment les systèmes d'assistance à l'exploitation ; l'accès à la conversion du numéro d'appel ou à des systèmes offrant des fonctionnalités équivalentes ; l'accès aux réseaux fixes et mobiles, notamment pour l'itinérance ; l'accès aux systèmes d'accès conditionnel pour les services de télévision numérique ; l'accès aux services de réseaux virtuels »³⁶. Cela impliquerait que les fournisseurs et/ou les distributeurs de contenus disposent d'un droit d'accès et d'interconnexion aux mêmes conditions que les fournisseurs de réseaux et services de communications électroniques actuellement, ce qui mettrait un terme au débat actuel visant à établir si les fournisseurs et/ou les distributeurs de contenus doivent ou non bénéficier directement des obligations imposées aux opérateurs de réseaux en matière d'accès. La modification proposée pour l'article 20 de la Directive « cadre » va dans le même sens, en spécifiant que les litiges entre « prestataires de services » tomberont dans le champ d'application des mécanismes de règlement des litiges des autorités de réglementation nationales à partir du moment où « l'une des parties est une entreprise fournissant des réseaux ou des services de communications électroniques » (c'est-à-dire que l'autre partie peut être un radiodiffuseur ou un prestataire de services propres à la société de l'information).

Il reste à voir si les propositions de la Commission seront retenues dans le texte définitif des directives. Le projet de rapport Trautmann proposait que les références aux services propres à la société de l'information et aux services de contenus radiodiffusés soient retirées de la définition de l'« accès », afin d'empêcher que cette réglementation ne couvre les aspects liés aux contenus, ce qui ouvrirait un champ de litiges beaucoup plus vaste (déjà couvert par la Directive SMAV et la Directive sur le commerce électronique)³⁷. Néanmoins, lors du vote de la Commission ITRE le 7 juillet 2008, le rapport Trautmann final comportait un amendement n°91 surprenant, qui suggérait de modifier le texte en « mise à disposition d'une autre entreprise, dans des conditions bien définies et de manière exclusive ou non exclusive, de ressources et/ou de services **en vue de la fourniture de services de communications électroniques ou de services informatiques ou de contenu radiodiffusé** »³⁸. A notre avis, cet amendement comporte une contradiction avec la définition des services de communications électroniques (laquelle exclut explicitement la fourniture de services de contenus ou la fourniture des services qui exercent un contrôle éditorial sur ledit contenu, *supra*) mais a été adopté néanmoins par le Parlement européen le 24 septembre 2008.

Les amendements proposés en matière de protection des consommateurs (par exemple la transparence pour les consommateurs, article 20 de la Directive « service universel ») restent limités aux situations contractuelles (c'est-à-dire aux relations entre les abonnés et les fournisseurs) et aux

36) Commission européenne, proposition pour une Directive du Parlement européen et du Conseil modifiant les Directives 2002/21/CE relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques, 2002/19/CE relative à l'accès aux réseaux et services de communications électroniques ainsi qu'à leur interconnexion, et 2002/20/CE relative à l'autorisation des réseaux et services de communications électroniques COM(2007)697 rev1, 13 novembre 2007, disponible sur : http://ec.europa.eu/information_society/policy/ecomm/library/proposals/index_en.htm

37) Parlement européen (Commission ITRE – rapporteur : Trautmann), Projet de rapport sur la proposition de directive du Parlement européen et du Conseil modifiant les Directives 2002/21/CE relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques, 2002/19/CE relative à l'accès aux réseaux et services de communications électroniques ainsi qu'à leur interconnexion, et 2002/20/CE relative à l'autorisation des réseaux et services de communications électroniques (COM(2007)0697 – C6-0427/2007 – 2007/0247(COD)) PE398.542v02-00, 23 avril 2008, amendement 60, p. 46 (voir également l'amendement 615 de Pleguezuelos, Amendements au projet de rapport Trautmann, vol. 5 (amendements 603-716), PE407.731v01-00, 10 juin 2008, p. 15), disponible sur : http://www.europarl.europa.eu/meetdocs/2004_2009/organes/itre/itre_20080707_1900.htm ou <http://www.europarl.europa.eu/activities/committees/publicationsCom.do?language=FR&body=ITRE>

38) Parlement européen (Commission ITRE – rapporteur : Trautmann), Rapport final sur la proposition de directive du Parlement européen et du Conseil modifiant la Directive 2002/21/CE relative à un cadre réglementaire commun pour les réseaux et services de communications électroniques, la Directive 2002/19/CE relative à l'accès aux réseaux et services de communications électroniques ainsi qu'à leur interconnexion, et la Directive 2002/20/CE relative à l'autorisation des réseaux et services de communications électroniques (COM(2007)0697 – C6-0427/2007 – 2007/0247(COD)), A6-0321/2008, PE398.542v03-00, 22 juillet 2008.

« entreprises fournissant des services et/ou des réseaux de communications électroniques » (et non aux fournisseurs de ressources associées).

À moins que certaines réformes radicales ne soient proposées par le Conseil, le nouveau dispositif sur les télécoms ne changera pas fondamentalement la position juridique des outils de recherche dans le cadre de la réglementation communautaire sur la transmission.

1.3. ...dans la Directive sur les services de médias audiovisuels

Le 11 décembre 2007, le législateur européen a enfin adopté la Directive sur les services de médias (Directive SMAV), qui modernise et remplace la Directive « Télévision sans frontières »³⁹. L'objectif de cette directive a toujours été la création d'un marché intérieur pour les services de télévision et, dans cette optique, l'instauration d'un minimum d'harmonisation dans certains domaines de la politique de radiodiffusion (y compris la protection des mineurs et de la dignité humaine, la diversité culturelle et le pluralisme dans les médias, l'accès à l'information et les communications commerciales).

La nouvelle directive ne couvre pas seulement les services de radiodiffusion classiques. Son champ d'application englobe à la fois les services de radiodiffusion linéaires, quelles que soient la technologie ou la plateforme utilisées, et les services de médias audiovisuels à la demande (tels que la vidéo à la demande). Ces services sont regroupés sous le terme global de « service de médias audiovisuels », défini comme « un service tel que défini aux articles 49 et 50 du Traité, qui relève de la responsabilité éditoriale d'un fournisseur de services de médias et dont l'objet principal est la fourniture de programmes dans le but d'informer, de divertir ou d'éduquer le grand public, par des réseaux de communications électroniques au sens de l'article 2, point a), de la Directive 2002/21/CE ».

Les différents éléments de cette définition, qui sont des critères cumulatifs⁴⁰, sont explicités dans plusieurs considérants. Ce serait sortir du cadre de cette contribution que de procéder à une analyse approfondie de ces critères⁴¹; nous nous contenterons donc d'examiner ceux qui sont pertinents pour les moteurs de recherche. Le considérant 18 spécifie que la définition du service de médias audiovisuels devrait couvrir les médias de masse en tant que moyens d'information, de divertissement et d'éducation du grand public, mais exclure tous les services dont la finalité principale n'est pas la fourniture de programmes, autrement dit les services dont le contenu audiovisuel est secondaire et ne constitue pas la finalité principale. « Tel est par exemple le cas des sites Web qui ne contiennent des éléments audiovisuels qu'à titre accessoire, comme des éléments graphiques animés, de brefs spots publicitaires ou des informations concernant un produit ou un service non audiovisuel »⁴². Le considérant 18 poursuit en ajoutant que « pour ces mêmes raisons », les jeux de hasard impliquant une mise correspondant à une valeur monétaire, y compris les loteries, les paris et les autres formes de jeux d'argent, de même que les jeux en ligne et les *moteurs de recherche*, devraient également être exclus du champ d'application de la présente directive.

Étant donné que dans la plupart des cas, les moteurs de recherche sur Internet ne remplissent pas le critère exigeant que la fourniture de programmes soit leur finalité principale, ils ne sont pas qualifiés pour être considérés comme des services de médias audiovisuels, même si, à l'occasion, ils présentent du matériel audiovisuel.

39) Directive 2007/65/CE du Parlement européen et du Conseil du 11 décembre 2007 modifiant la Directive 89/552/CEE du Parlement européen et du Conseil du 3 octobre 1989 visant à la coordination de certaines dispositions législatives, réglementaires et administratives des États membres relatives à la fourniture de services de médias audiovisuels (Directive « Services de médias audiovisuels ») [2007] JO L332/27.

40) Cf. considérant 25 : « Toutes les caractéristiques d'un service de médias audiovisuels spécifiées dans sa définition et exposées dans les considérants 16 à 23 doivent être réunies simultanément .»

41) Voir Remy Chavannes et Oliver Castendyk, Alphen a/d Rijn, 2008. Chapitre I. Définitions. Dans *European Media Law*, édité par Oliver Castendyk, Egbert Dommering et Alexander Scheuer, 813-819. Austin/Boston/Chicago/New York/The Netherlands: *Kluwer Law International* ; Peggy Valcke et Eva Lievens Eva. 2008. *Rethinking European broadcasting regulation - Unravelling Europe's policy for the digital media landscape: critical analysis of the Audiovisual Media Services Directive*. Dans *Rethinking European Media and Communications Policies*, édité par Caroline Pauwels et Harri Kalimo. Bruxelles : VUB Press (prochainement).

42) Par exemple le site d'une agence de voyage ou d'un constructeur automobile sur lequel on peut voir des clips présentant des sites de villégiature ou les derniers modèles automobiles.

Si les moteurs de recherche sont destinés à présenter des clips vidéo, comme Google Vidéo, certains analystes avancent qu'ils pourraient être considérés comme des services de médias audiovisuels, au même titre qu'un catalogue de programmes⁴³. De leur point de vue, le tout est de savoir si une telle offre présente un caractère de « type télévisuel » suffisant (cf. considérant 17) pour que sa forme et son contenu soient « comparables à ceux de la radiodiffusion télévisuelle », comme le formule l'article 1, b) de la directive (contenant la définition d'un « programme ») et, par conséquent, de savoir sous quelle forme le moteur de recherche est présenté au public. À cet égard, ils se réfèrent à une version de YouTube (offrant indéniablement des fonctions de recherche) accessible sur AppleTV, qui constitue un exemple parfait de service entrant dans cette catégorie.

Pour ma part, cependant, je pense que l'absence de responsabilité éditoriale, au sens visé par la directive, ne permet pas de considérer un moteur de recherche audiovisuel comme un service de médias audiovisuels. La notion de « responsabilité éditoriale » est devenue, sous l'impulsion du Parlement et du Conseil, un élément clef de la définition du service de média audiovisuel⁴⁴. Elle est définie à l'article 1 c) de la Directive SMAV comme « l'exercice d'un contrôle effectif tant sur la sélection des programmes que sur leur organisation, soit sur une grille chronologique, dans le cas d'émissions télévisées, soit sur un catalogue, dans le cas de services de médias audiovisuels à la demande. La responsabilité éditoriale n'a pas nécessairement pour corollaire une responsabilité juridique quelconque en vertu du droit national à l'égard du contenu ou des services fournis ».

On peut difficilement prétendre que la fourniture automatisée de résultats basés sur des algorithmes soit comparable au contrôle effectif sur la sélection et sur l'organisation des programmes sur une grille chronologique ou sur un catalogue par un radiodiffuseur ou un fournisseur de vidéos à la demande. Même si des moteurs de recherche spécialisés pour les vidéos ou les sites de vidéos tels que YouTube ont adopté une pratique consistant à retirer les vidéos classées impropres par les internautes membres de la communauté ; cette intervention se fait a posteriori, en réaction aux plaintes des utilisateurs⁴⁵. En outre, le contrôle éditorial est « réparti », comme le qualifie Van Hoboken, c'est-à-dire qu'il est partagé entre les fournisseurs des informations et les internautes eux-mêmes, qui marquent et classent les contenus qu'ils placent et consultent en ligne⁴⁶.

Que se passerait-il si la sélection et la présentation du matériel audiovisuel étaient réalisées par des équipes humaines disposant de centres éditoriaux ? Dans ce cas, on se rapprocherait de la notion de guides électroniques de programmes tels qu'ils sont mentionnés par le considérant 22 : « Si le principal objectif d'un service de média audiovisuel est la fourniture de programmes, la définition d'un tel service devrait également s'appliquer au contenu fondé sur le texte qui accompagne de tels programmes, comme les services de sous-titrage et les guides électroniques de programmes. Les services textuels autonomes ne devraient pas relever de la présente directive, ce qui ne devrait pas porter atteinte à la liberté des États membres de réglementer ces services au niveau national, conformément au traité. »

Autrement dit, les guides électroniques de programmes ne sont pas considérés comme un service de médias audiovisuels en tant que tels (puisque leur finalité principale n'est pas la fourniture de programmes, mais la présentation de ces programmes associée à des informations et fonctionnalités supplémentaires telles que zapping, enregistrement, etc.), mais s'ils accompagnent un service de médias audiovisuels, ils entrent cependant dans le champ d'application de la directive. Le guide électronique de programmes est alors considéré comme faisant partie des prestations du fournisseur de services des médias audiovisuels et devrait, à ce titre, être également assujéti aux règles de la directive (notamment celles concernant la protection des mineurs et de la dignité humaine ou, *mutatis mutandis*, les règles relatives aux communications audiovisuelles commerciales).

A contrario, si le guide électronique de programmes est proposé comme un service autonome (c'est-à-dire non compris dans un service de médias audiovisuels), par exemple par un opérateur de câble ou de satellite, les dispositions de la directive ne sont plus applicables.

43) Remy Chavannes et Oliver Castendyk. 2008. *op.cit.*, 814.

44) Voir Wolfgang Schulz et Stefan Heilmann, 2008. Dans *IRIS Spécial : La responsabilité éditoriale*, édité par Susanne Nikoltchev, Strasbourg, Observatoire européen de l'audiovisuel.

45) Voir les règles de YouTube Community sur http://www.youtube.com/t/community_guidelines

46) Joris Van Hoboken, 2008. *Legal Space for Innovative Ordering – The need for more clarity regarding the liability of selection intermediaries for third party information*. Rapport présenté à la conférence EuroCPR2008, Séville, mars 2008, p. 4.

1.4. ...dans la Directive sur le commerce électronique

La Directive sur le commerce électronique⁴⁷ adoptée en 2000, vise à instaurer un cadre juridique de base pour le commerce électronique dans le marché intérieur et favoriser le développement des services en ligne transfrontaliers. Elle s'applique aux services de la société de l'information, définis comme « tout service presté normalement contre rémunération, à distance par voie électronique et à la demande individuelle d'un destinataire de services. »⁴⁸ Cette notion couvre une grande variété de services fournis en ligne, depuis les journaux en ligne, les services d'actualité spécialisés (tels que les informations financières ou économiques) et la vente en ligne de divers produits et services (tels que livres, DVD, matériel et logiciels informatiques, services de voyages, pharmaceutiques, etc.) jusqu'à la fourniture en ligne de services financiers (gestion de compte en ligne et placements en ligne).

Le considérant 18 de la directive précise : « Les services de la société de l'information ne se limitent pas exclusivement aux services donnant lieu à la conclusion de contrats en ligne, mais, dans la mesure où ils représentent une activité économique, ils s'étendent à des services qui ne sont pas rémunérés par ceux qui les reçoivent, tels que les services qui fournissent des informations en ligne ou des communications commerciales, ou ceux qui fournissent des outils permettant la recherche, l'accès et la récupération des données. » Autrement dit, les moteurs de recherche sont, en principe, inclus dans le champ d'application de la directive.

Mais les règles pertinentes restent limitées. Tout d'abord, parce que les domaines coordonnés par la directive et l'harmonisation ainsi obtenue restent d'une portée limitée, et ensuite parce que les règles qui semblent les plus pertinentes pour les moteurs de recherche, notamment les règles concernant la responsabilité des prestataires de services intermédiaires, ne s'appliquent pas à eux (mais aux seules activités de routage, d'hébergement et de mise en mémoire cache). Nous n'allons pas nous attarder sur ce point, qui est analysé de façon approfondie dans le rapport de Joris van Hoboken⁴⁹.

1.5. Conclusion intermédiaire

Nous avons vu dans les paragraphes précédents que le cadre réglementaire actuel de l'UE ne traite que de façon très fragmentaire les outils de recherche audiovisuels. Il existe des dispositions limitées, tant dans les Directives sur les communications électroniques que dans la Directive sur les services de médias audiovisuels, concernant les guides électroniques de programmes (que nous considérons ici comme des « outils de recherche audiovisuels mis en forme », dans un environnement *walled-garden* de la radiodiffusion numérique. En revanche, les outils de recherche en ligne, alors même qu'ils sont inclus dans le champ d'application de la Directive sur le commerce électronique, ne sont pas couverts par les dispositions refuges de la Directive sur le commerce électronique. Autrement dit, la situation concernant le statut juridique des outils de recherche n'a pratiquement pas évolué depuis l'analyse de Nico van Eijk dans IRIS *plus* 2006-2 (« Recherche et tu trouveras ? »).

2. Réflexions préliminaires sur la régulation de la fonction de recherche audiovisuelle

2.1. Est-il besoin de réguler ?

La question de la nécessité ou non de réguler est fondamentale. Nous devons avoir de bons arguments pour réguler des fonctions pour lesquelles nous n'avons jamais, auparavant, promulgué de règles sectorielles dans l'environnement analogique et imprimé ; il suffit de considérer les guides de présenta-

47) Directive 2000/31/CE du Parlement européen et du Conseil, du 8 juin 2000, relative à certains aspects juridiques des services de la société de l'information, et notamment du commerce électronique, dans le marché intérieur (« Directive sur le commerce électronique ») [2000] JO L 178/1.

48) Article 1 (a) de la Directive 98/34/CE du 22 juin 1998 prévoyant une procédure d'information dans le domaine des normes et réglementations techniques [1998] JO L 204/37 dans sa version modifiée par la Directive 98/48/CE du 20 juillet 1998 [1998] JO L217/17.

49) Voir Joris van Hoboken, 2008. *Legal Space for Innovative Ordering – The need for more clarity regarding the liability of selection intermediaries for third party information*. Rapport présenté à la conférence EuroCPR2008, Séville, mars 2008. Voir également l'article de Joris van Hoboken dans cette publication.

tion des programmes télévisés, les librairies, les kiosques à journaux, les bibliothèques : depuis de nombreuses années, tous ces dispositifs servent à sélectionner des contenus et en faciliter l'accès, sans que quiconque ne défende l'idée qu'il soit nécessaire de leur imposer des règles spécifiques de neutralité ou autres. En quoi l'environnement numérique en ligne est-il si différent, hormis le fait que, pour l'instant, il nous est moins familier et que les fournisseurs des services de recherche peuvent être implantés pratiquement à n'importe quel endroit de la planète ?⁵⁰ Dans le cadre des moteurs de recherche en ligne basés sur du texte, la manipulation (cachée) des résultats de la recherche et la prédominance d'un opérateur (Google) ont été citées comme des motifs pouvant justifier la mise en place d'une intervention réglementaire⁵¹. Mais de telles raisons n'ont pas encore été mises en avant en ce qui concerne les outils de recherche audiovisuels.

2.2. Sur qui / sur quoi doit porter la régulation ?

La deuxième question est : sur qui / sur quoi doit porter la régulation ? Faut-il se limiter aux outils de recherche audiovisuels impliquant une certaine forme de contrôle éditorial (tel que la sélection préalable des chaînes ou des programmes qui seront répertoriés et présentés sous une forme interrogeable), ou englober également les services de recherche automatisés (avec, au mieux, retrait de certaines références a posteriori, sur plainte des internautes) ? On serait enclin à pencher pour la première solution, puisque dans ce cas, les outils de recherche impliquent un contrôle plus important de la part du fournisseur. Mais, ironie de la situation, cela risque de créer des effets pervers : plus on contrôle, plus on est régulé, donc allégeons le contrôle...

2.3. Pourquoi voulons-nous réguler (et comment) ?

Enfin, troisième question : quels sont les objectifs exigeant une intervention réglementaire (et comment les réaliser au mieux ?). Les préoccupations concernant la domination et la mainmise sur le marché peuvent justifier une intervention, mais les premiers outils permettant d'assurer une concurrence libre et loyale sont les règles générales sur la concurrence (en premier lieu, dans notre cas, l'interdiction visée à l'article 82 du Traité CE d'abuser d'une position dominante sur le marché)⁵². Seuls les secteurs autrefois marqués par des monopoles ont vu l'introduction d'outils sectoriels pour appliquer a priori des mesures aux entreprises ayant une certaine position sur le marché (comme le régime de puissance significative sur le marché des télécommunications). Le secteur de la télévision numérique constituait un cas particulier, avec des obligations d'accès pour les fournisseurs de services à accès conditionnel, indépendamment de leur position sur le marché. La Directive « accès » actuelle a hérité de ce régime et instauré des règles comparables (mais non similaires) pour les « nouveaux » goulots d'étranglement numériques, tels que les interfaces de programme d'application et les guides électroniques de programmes. Comme nous l'avons déjà vu, ces règles concernent exclusivement les aspects liés à la transmission de ces ressources et sont muettes sur les questions de présentation ou de contenu.

En ce qui concerne les objectifs culturels (tels que le pluralisme des médias), une intervention pourrait s'avérer nécessaire si certains contenus considérés d'intérêt public n'apparaissaient plus de façon visible dans les listes de recherche. Par conséquent la solution suggérée par Cullen dans son étude de 2006, qui consiste à spécifier que l'obligation de distribution (*must-carry*) doit garantir l'accès au réseau de transmission et à toutes les ressources associées (qu'il considère comme des fonctions de distribution, donc couvrant non seulement les services à accès conditionnel (CAS), API et EPG, mais aussi les systèmes de gestion des droits numériques (DRM), les décodeurs, les multiplexes et les moteurs de recherche⁵³) semble être une piste intéressante, bien que ce ne soit qu'une première étape.

50) Néanmoins, s'ils veulent cibler un public spécifiquement défini au niveau géographique, ils devront, à plus ou moins long terme, implanter au moins une section ou une représentation locale pour rester en phase avec la culture et les habitudes du public et entretenir des liens étroits avec les annonceurs de la région.

51) Van Eijk 2006, *op. cit.* p. 6 et Nico van Eijk, 2007. *Search Engines, the new bottleneck for content access*. Rapport présenté lors de l'*International Telecommunications Society 19th European Regional Conference*, Istanbul, du 2 au 5 septembre 2007, p. 20, disponible sur : http://www.ivir.nl/publications/vaneijk/Paper_SearchEngines_ITS_2007.pdf

52) Des règles de contrôle des fusions pourraient également apparaître si les principaux exploitants de moteurs de recherche entamaient un processus d'intégration verticale ou horizontale par le biais d'alliances et de reprises, ou si l'interdiction des ententes de cartel et des pratiques concertées s'appliquaient aux ententes (exclusivement) verticales entre fournisseurs de contenus et exploitants de moteurs de recherche.

53) Cullen International, *op. cit.*, note 34, p. 120 et 125.

En effet, dans un domaine où les informations foisonnent, l'important n'est pas d'être simplement répertorié, mais d'être répertorié *en bonne place*. À titre d'exemple de mesures poursuivant cet objectif⁵⁴, on peut citer les dispositions allemandes relatives aux navigateurs dans la *Satzung digitaler Zugang*, qui comportent l'obligation d'assurer une présentation « équilibrée » des programmes proposés sur les chaînes publiques et privées.

La protection des consommateurs peut constituer un autre objectif pertinent dans le domaine des outils de recherche audiovisuels. Dans ce cadre, on pourrait envisager des règles visant à instaurer la transparence des méthodes de classement (et des « liens sponsorisés ») ainsi que des dispositions régissant le retrait (et/ou la mise en garde) des résultats faisant référence à des contenus illicites ou préjudiciables (telles que la définition des modalités et procédures de ces retraits et mises en garde). À cet égard, on peut citer le code de bonnes pratiques des moteurs de recherche en Allemagne, qui est présenté dans la contribution de Wolfgang Schulz⁵⁵.

La protection de la vie privée pourrait être une étoile montante au firmament des objectifs, en justifiant une intervention (plus) règlementaire. Les préoccupations concernant la vie privée grandissent au vu des technologies émergentes qui sont conçues pour suivre et conserver l'historique des activités des utilisateurs en ligne ou sur les plateformes de télévision afin de personnaliser les services. Comme l'a démontré Michael Zimmer⁵⁶, des douzaines de produits et de services sont reliés par les cookies et les comptes des utilisateurs, et les gens laissent derrière eux un nombre croissant « d'empreintes digitales » sur Internet (tels que contacts, blogs, stocks, rendez-vous, e-mails, amis, dossiers informatiques, groupes de discussion, adresses URL). En outre, ils perdent le contrôle sur ce que Michael Zimmer appelle « leur portrait en ligne » ; en effet, il se peut qu'ils n'aient jamais connaissance (sans parler de la possibilité de donner leur accord) du prélèvement et de la publication en ligne par des tiers de certaines informations à caractère personnel les concernant (y compris de matériel audiovisuel tel que photos ou vidéos). À cet égard, nous aimerions attirer l'attention du lecteur sur l'Avis 1/2008 sur les aspects de la protection des données liés aux moteurs de recherche émis récemment par le groupe de travail Article 29 (sous la pression de l'inquiétude croissante des utilisateurs européens quant aux éventuelles atteintes à la vie privée). Cet avis examine l'activité des moteurs de recherche et le stockage des données au vu des directives européennes concernant la protection des informations personnelles et la conservation des données⁵⁷. Il y est expliqué que les fournisseurs de moteurs de recherche, en fonction de leurs activités précises, seront considérés à maints égards comme des contrôleurs des données à caractère personnel et devront se conformer aux règles et aux principes en vigueur. Cela englobe l'obligation d'effacer ou d'assurer l'anonymat des données à caractère personnel dès qu'elles ne servent plus les finalités pour lesquelles elles ont été collectées, ainsi que la nécessité pour les fournisseurs de moteurs de recherche spécialisés dans la création d'opération à valeur ajoutée telles que les profils de personnes physiques (« *people search engines* ») et les logiciels de reconnaissance faciale sur images et contenus audiovisuels de veiller à ce que leur traitement soit légitime, notamment par le consentement des personnes concernées, et de se conformer par ailleurs à toutes les exigences de la Directive 95/46/CE sur la protection des données (telles que l'obligation de garantir la qualité des données et un traitement équitable).

3. Conclusion

Actuellement, tant les guides électroniques de programmes que les moteurs de recherche sortent largement du champ d'application des cadres règlementaires spécifiques aux secteurs des médias et des

54) *Satzung über die Zugangsfreiheit zu digitalen Diensten gemäß § 53 Abs. 6 Rundfunkstaatsvertrag, § 13 (3)* (Statut relatif à la liberté d'accès aux services numériques selon l'article 53, paragraphe 6 *Rundfunkstaatsvertrag*, article 13 (3)) : « *Auf das öffentlich-rechtliche und private Programmangebot muss gleichgewichtig hingewiesen werden.* » (adopté par la *Direktorenkonferenz der Landesmedienanstalten (DLM)* le 13 décembre 2005, http://www.lfk.de/gesetzundrichtlinien/rundfunkstaatsvertrag/download/Satzung_digitale_Dienste.pdf).

55) Voir l'article de Wolfgang Schulz dans cette publication.

56) Voir l'article de Michael Zimmer dans cette publication.

57) Groupe de travail Article 29, Avis 1/2008 sur les aspects de la protection des données liés aux moteurs de recherche, WP 148, 4 avril 2008. Le groupe de travail Article 29 est un organe consultatif de l'Union européenne composé de représentants des instances de protection des données des États membres. Les avis du groupe de travail Article 29 ne sont pas juridiquement contraignants mais font souvent office de « norme UE », ce qui leur vaut d'exercer une forte influence sur l'élaboration de la politique UE en matière de vie privée et sa mise en application par les États membres.

télécommunications. Néanmoins, nous sommes d'avis que la plupart des objectifs de réglementation, qui répondent aux préoccupations croissantes concernant les outils de recherche, peuvent fort bien être atteints en appliquant les règles et les principes généraux en vigueur, tels que les règles sur la concurrence (pour répondre aux préoccupations concernant l'hégémonie sur le marché), la législation sur la protection des données (en réponse à l'inquiétude concernant les atteintes à la vie privée) ou les règles sur les pratiques commerciales déloyales et la publicité (pour renforcer la transparence au niveau de la « manipulation » des résultats). La solution consiste peut-être davantage à œuvrer à une plus grande cohérence des règles existantes qu'à promulguer de nouvelles règles.

Cependant, nous convenons avec les observateurs soucieux⁵⁸ que les développements doivent être surveillés de près, puisque la fonction de recherche devient effectivement un élément crucial à l'ère du foisonnement des informations (peut-être même plus que la fourniture de contenus elle-même, à présent que nous assistons à une baisse sans précédent des obstacles au statut de fournisseur de contenu).

Dans le cadre audiovisuel, la fonction de recherche est souvent associée à d'autres fonctions de distribution (telles que la fourniture d'une plateforme d'hébergement ou d'échange de contenus audiovisuels), ce qui nous amène à nous demander si la portée de la réglementation ne devrait pas dépasser la fonction de recherche en tant que telle. Dans certains États membres de l'UE⁵⁹ et au Japon⁶⁰, les législateurs ont introduit ou sont en passe d'introduire un schéma de réglementation à trois niveaux pour le secteur des info-communications, en faisant la distinction entre les contenus, la distribution et la transmission⁶¹. Par conséquent, les fournisseurs des « fonctions intermédiaires » (c'est-à-dire les intermédiaires entre la configuration et la fourniture des contenus, d'une part et la fourniture des réseaux et services de transmission par le biais desquels transitent les contenus d'autre part) relèveraient d'un cadre réglementaire commun. Si tel était le cas au niveau de l'UE, cela permettrait au législateur européen de mettre en place des règles et des principes de base applicables (tels que l'ouverture et la loyauté), tout en laissant suffisamment de place pour affiner ces règles de base par le biais, essentiellement, d'outils de corégulation ou d'autorégulation. En définitive, trouver le parfait équilibre entre des objectifs conflictuels de réglementation et/ou les droits et libertés peut s'avérer un processus de longue haleine. Où finit la liberté d'expression de l'opérateur de recherche et, partant, sa liberté de pratiquer une sélection active, et où commence l'obligation de non-discrimination ? Comment peut-on concilier la responsabilité de l'opérateur de recherche et l'extension de la sélection active ? Nul doute que ces questions vont encore nous occuper pendant de longues années.

Le risque existe d'une régulation excessive et trop précoce (ce qui peut entraîner de fausses mesures d'incitation et étouffer l'innovation) ; les promesses des nouvelles technologies ne devraient être tempérées que lorsque la menace devient manifeste et sérieuse (sans toutefois nous laisser prendre au marché faustien mentionné par Michael Zimmer⁶²). Peut-être viendra-t-il un jour où nous maudirons les outils de recherche (lorsqu'ils commenceront à nous connaître mieux que nous-mêmes) mais pour l'instant, je pense qu'on peut encore les considérer comme des outils extrêmement utiles pour trouver

58) Voir à cet égard le commentaire critique sur le manque de transparence des moteurs de recherche sur Internet et les manipulations cachées des résultats : Marcel Machill, Christoph Neuberger et Friedemann Schindler, 2003. *Transparency on the Net: functions and deficiencies of Internet search engines*. 5 *Info* 1, p. 52-74.

59) Notamment la Belgique (la communauté francophone a introduit un système à trois niveaux dans la loi sur la radiodiffusion de 2003, comme nous l'avons expliqué dans notre précédent rapport « L'avenir de la notion d'obligation de distribution », 2005, dans *IRIS Spécial : Avoir ou ne pas avoir – les règles du must-carry*, édité par Susanne Nikoltchev, Strasbourg, Observatoire européen de l'audiovisuel. La communauté flamande adoptera un système similaire à trois niveaux dans sa nouvelle loi sur les médias qui vise à transposer la Directive SMAV) ; voir également en France la loi n°86-1067 du 30 septembre 1986 relative à la liberté de communication (« loi Léotard ») qui fait état du concept de « distributeur » (en particulier dans le cadre de la radiodiffusion numérique terrestre).

60) Dans son rapport présenté à la 17^e conférence biennale de l'International Telecommunications Society qui s'est tenue à Montréal du 24 au 27 juin 2008, le professeur Minoru Sugaya (Institut des médias et de la communication de l'Université de Keio) analyse les réformes actuelles de la réglementation en expliquant comment le Japon entend intégrer le cadre existant en matière de (télé)communications et de radiodiffusion au sein d'un dispositif unique de « loi sur l'information et les communications », structuré en fonction des « contenus », « plateformes » (couche intermédiaire) et « infrastructure de transmission » (<http://www.canavents.com/its2008/plenary3/13.pdf> ; pour de plus amples informations sur les recommandations du groupe de travail sur un Système juridique global pour la radiodiffusion et les communications auquel se réfère le rapport de Minoru Sugaya, disponible sur : http://www.soumu.go.jp/joho_tsusin/eng/Releases/NewsLetter/Vol18/Vol18_21/Vol18_21.html).

61) Ou à quatre niveaux, en fait, si l'on considère l'équipement terminal comme un quatrième niveau à réguler.

62) Voir l'article de Michael Zimmer dans cette publication.

des informations pertinentes. Même en admettant qu'il y a un grand nombre d'informations que nous ne trouvons pas à l'aide des outils de recherche les plus connus parce qu'ils classent d'autres réponses en priorité pour des raisons commerciales ou parce qu'ils filtrent certains éléments, n'y aurait-il pas *beaucoup plus* d'informations que ne trouverions pas *sans* eux (non, détrompez-vous, je ne suis pas sponsorisée par eux pour tenir ces propos...)

Il est certain que les prestataires de services de recherche doivent prendre conscience du rôle croissant qu'ils jouent dans la structuration de la sphère publique et de leur impact grandissant sur la société. Par conséquent, ils devraient se préparer à assumer leurs responsabilités sociales et, dans le cas contraire, les pouvoirs publics peuvent être amenés à intervenir pour « brider » leurs pouvoirs privés. À cet égard, notre rôle, en tant que chercheurs, consiste à développer et poursuivre le débat public sur ces questions, puisqu'on ne peut pas laisser aux fournisseurs le soin de mettre en place eux-mêmes les normes appropriées ; en effet, ils ne sont pas censés penser et agir dans l'intérêt général. Dans ce cadre, les initiatives visant à approfondir les analyses et les recherches sont extrêmement utiles. Outils de recherche, vous êtes prévenus : si vous nous surveillez, sachez que nous vous surveillons également !

Réguler les moteurs de recherche ? Le recours à l'autorégulation et à la corégulation dans le domaine de la recherche sur Internet

Wolfgang Schulz*
Hans-Bredow-Institut

Introduction

De nos jours – en Allemagne, tout du moins –, lorsque des dirigeants de médias venus de secteurs différents se réunissent, c'est généralement pour se chamailler au sujet des difficultés économiques et de la concurrence acharnée qu'ils rencontrent. Mais il est un sujet commun qui permet à tous de renouer le dialogue : le danger que représente Google, qui menace de grignoter le marché de la publicité.

De ce fait, la régulation des moteurs de recherche, leur influence potentielle sur la formation de l'opinion publique, leur capacité de discrimination, ainsi que d'autres questions de nature réglementaire, sont passées au premier plan dans l'agenda politique. Toutefois, l'appréhension de ce phénomène reste encore tâtonnante. Le présent article propose une définition du concept d'« intermédiaire en information » (moteurs de recherche, par exemple), identifie les objectifs de politique publique qui pourraient se trouver menacés, et s'interroge sur la pertinence des concepts de corégulation et d'autorégulation face aux défis de régulation qui se posent.

1. Le concept d'intermédiaire en information

Les modèles de régulation des médias s'appuient, implicitement ou explicitement, sur un certain nombre de hypothèses concernant le fonctionnement des communications personnelles et publiques. La protection accordée spécifiquement aux médias de masse par l'article 10 de la Convention européenne des Droits de l'Homme (CEDH), ou par l'article 5 alinéa 1 de la Loi fondamentale (Grundgesetz – GG – Constitution allemande) repose sur le postulat suivant : dans une société démocratique, les médias de masse remplissent une fonction particulière en matière de communication publique et personnelle¹. La Cour constitutionnelle allemande, qui a explicitement souligné le rôle des médias de masse, opère une distinction entre la protection de l'expression, c'est-à-dire du contenu effectif de la communication (qu'elle soit émise sous forme de communication personnelle ou publiée par un média de masse), et la protection des médias de masse (radiodiffusion, presse et cinéma sont mentionnés dans la constitution, mais ce terme peut également recouvrir les nouveaux médias de masse)². De ce

* Wolfgang Schulz est directeur du Hans-Bredow-Institut et maître de conférence à la Faculté de droit de l'Université de Hambourg.

1) Hoffmann-Riem/Schulz dans : *Politische Kommunikation in der demokratischen Gesellschaft*, Wiesbaden 1998, 154 (p. 159 et suivantes) (en allemand uniquement) ; Hoffmann-Riem, *Regulating Media*, New York 1996, 119.

2) Voir décisions de la Cour constitutionnelle allemande (en allemand uniquement) : BVerfGE 73, 118 (158) ; 74, 297 (324) ; 83, 238 (298) ; 90, 60 ; voir également la décision la plus récente : BVerfG, 1 BvR 2270/05, 11 septembre 2007 (uniquement en allemand), disponible sur : http://www.bverfg.de/entscheidungen/rs20070911_1bvr227005.html

préalable, la Cour constitutionnelle fédérale déduit les formes spécifiques de protection constitutionnelle contre l'ingérence de l'Etat, ainsi que les obligations constitutionnelles incombant au législateur, et visant à ce que le pouvoir de formation de l'opinion publique (exercé en principe par ceux qui contrôlent les médias de masse) fasse l'objet d'un contrôle effectif.

Ce rôle particulier de la communication par les médias de masse est une hypothèse qui reste dans la droite lignée des découvertes des sciences de la communication contemporaines. Pourtant, un changement structurel s'est produit dans les réseaux publics de communication. Tandis que les médias de masse se caractérisent par l'exercice d'un contrôle sur ce qui est publié, mais aussi par une présélection rédactionnelle, on a vu apparaître un nombre croissant de formes de communication qui ménagent une grande importance structurelle à d'autres types de prestataires de services. En d'autres termes, lorsqu'il y a ingérence de l'Etat dans ces services, elle ne touche pas un seul acte de communication, mais la structure générale des communications publiques. Ces autres types de prestataires de services, nous les désignerons par l'expression « prestataires intermédiaires de services ». Les sciences de la communication sont encore en train de caractériser les différents types de services en ligne qui pourraient, à l'avenir, constituer la base du débat juridique. Certains types de plateformes, comme les forums Internet, peuvent être considérés comme des « intermédiaires en information ».

L'importance du rôle des moteurs de recherche a été démontrée dans plusieurs études³. Il est impossible de savoir si l'affirmation selon laquelle l'indexation par un moteur de recherche est un préalable nécessaire pour « exister » sur Internet⁴ reflète la réalité ou constitue une exagération. Toutefois, il est évident que les moteurs de recherche remplissent une fonction fondamentale permettant de rendre accessible l'information sur Internet. Même si la dépendance des utilisateurs vis-à-vis des moteurs de recherche diminue à mesure qu'ils deviennent de plus en plus experts dans l'utilisation d'Internet, cette fonction de base demeure importante.

On pourrait croire à première vue que les moteurs de recherche, intermédiaires de communication spécialisés, n'ont rien en commun avec les prestataires de services classiques en matière de médias. Il n'y a pas ici de sélection préalable, qu'elle soit journalistique ou rédactionnelle ; le moteur de recherche n'affiche que ce que l'utilisateur a demandé dans sa requête. Schönbach a souligné cette différence en qualifiant les médias traditionnels de « médias d'affichage » (*display media*). Là, c'est la présélection qui doit surprendre l'utilisateur : face à un média classique, l'utilisateur s'attend à ce que quelqu'un d'autre choisisse le contenu qui lui sera présenté en fonction d'une norme professionnelle donnée, et que ce contenu le surprenne. De l'autre côté, nous trouvons ce que l'on appelle les « outils de recherche », dont les moteurs de recherche sont un exemple paradigmatique et de premier plan⁵. Toutefois, à y regarder de plus près, cette classification des moteurs de recherche ne correspond pas tout à fait aux critères applicables, notamment si l'on songe que certains des algorithmes utilisés par les moteurs de recherche s'efforcent de cibler les pages les plus à jour, ou que les moteurs de recherche utilisent des indicateurs de hiérarchie, qui constituent aussi des choix journalistiques et rédactionnels. Il reste que la fonction principale, qui motive l'utilisation des moteurs de recherche, est celle qui consiste à détecter une information particulière dont l'utilisateur connaît l'existence, mais pas l'emplacement exact.

Le rôle spécifique des intermédiaires en information dans les communications personnelles et publiques peut faire apparaître des risques caractéristiques. Néanmoins, la fonction particulière que remplissent les moteurs de recherche dans les communications personnelles et publiques peut elle aussi nécessiter une protection spécifique. Elle peut ainsi requérir la mise en œuvre de garde-fous ciblés contre la régulation,

3) Voir Van Eijk, Nico. Moteurs de recherche : Cherche et tu trouveras ? Positionnement juridique des moteurs de recherche. IRIS plus 2006-7 ; Strasbourg, Observatoire européen de l'audiovisuel ; disponible sur : http://www.obs.coe.int/oea_publ/iris/iris_plus/iplus2_2006.pdf ; Philippi, *Internet Suchmaschinen: Suchmethodik – Erlösmodelle – Preisstrategien*, (2006) ; Nitsche, *Suchmaschinen im Internet - Bedeutung, Trends, Chancen und Entwicklungen* (2005) ; Machill/ Neuberger/ Schweiger/ Wirth, *Wegweiser im Netz - Qualität und Nutzung von Suchmaschinen* (2003) ; Introna/Nissbaum, *Shaping the Web: Why the Politics of Search Engines Matters, The Information Society* 171 (2000) ; Maurer et al., *Report on dangers and opportunities posed by large search engines, particularly Google* (2008), disponible sur : <http://www.google-watch.org/gpower.pdf>

4) Voir Introna/Nissbaum, *Shaping the Web: Why the Politics of Search Engines Matters, The Information Society* 171 (2000).

5) Voir Schönbach, *Das Eigene im Fremden. Zuverlässige Überraschung – eine wesentliche Medienfunktion?*, *Publizistik*, 2005 (50) : 344-352.

plutôt que d'appeler une action législative. Ce choix dépend de la structure du cadre réglementaire donné, question qui ne peut être abordée ici dans le détail. Dans ce contexte, il est indispensable de garder à l'esprit les droits fondamentaux, et notamment la garantie de la liberté de parole⁶.

2. Les risques engendrés par les moteurs de recherche

Cet article s'intéresse aux risques que présentent les moteurs de recherche en ligne et aux approches réglementaires envisageables pour y faire face.

Les risques engendrés par les moteurs de recherche sur Internet sont multiples :

- Accès à des contenus préjudiciables
- Accès à des contenus illégaux

- Exploitation d'œuvres protégées
- Exploitation de données personnelles

- Discrimination de certains contenus
- Consommateurs induits en erreur
- Influence sur la formation de l'opinion

- Distorsion de la concurrence, notamment :
transfert du pouvoir de marché à d'autres marchés (publicité, par exemple)

- Fragmentation de l'espace public

L'un des inconvénients inhérents au rôle clé que jouent les moteurs de recherche en tant qu'accès (voire, contrôleurs d'accès) à Internet, est que toute question juridique qui touche Internet risque tôt ou tard de les concerner.

Le premier ensemble de risques engendrés par les moteurs de recherche tient au fait qu'ils rendent accessibles (plus facilement) des contenus préjudiciables ou illégaux. La définition d'un contenu préjudiciable ou illégal varie d'un pays à l'autre. La pédopornographie⁷ est jugée illégale sans exception dans tous les pays ; dans la plupart d'entre eux, il existe en outre des restrictions particulières concernant l'accès des mineurs aux contenus susceptibles d'entraver leur développement⁸.

Un autre ensemble de risques peut apparaître du fait que les fournisseurs de moteurs de recherche ont la possibilité de sélectionner les pages qu'ils incluront dans leur index, d'en supprimer certaines et de présenter à leur guise la liste de résultats. Cela peut conduire à discriminer certains contenus, mais aussi, dans le cas des communications commerciales, à induire en erreur les consommateurs, et,

6) Pour des moteurs de recherche liés au droit constitutionnel allemand, voir Schulz, *Von der Medienfreiheit zum Grundrechtsschutz für Intermediäre?* CR 2008, p. 470 et seq.

7) Voir par exemple l'alinéa 4 de la JMStV (Allemagne) ; Loi fédérale 18 U.S.C. alinéa 2252 (Etats-Unis) ; pour l'Europe, voir également http://www.saferinternet.org/ww/en/pub/insafe/safety_issues/faqs/pornography.htm ; l'harmonisation menée par le Conseil de l'Europe via la Convention sur la cybercriminalité (STE n° 185).

8) Voir par exemple : *Awareness and Harmful Media Content in the Digital Age*, Ulla Carlsson (Ed.) (2006), disponible sur : <http://unesdoc.unesco.org/images/0014/001469/146955E.pdf>

en règle générale, à peser sur le processus de formation de l'opinion. Pendant la guerre en Iraq, des rumeurs affirmaient que certains moteurs de recherche supprimaient de leur index les photos montrant des soldats américains en train de torturer des prisonniers⁹. Etant donné que Google fournit plus de 80 % des résultats des recherches menées en Allemagne, on peut affirmer sans exagérer que les premiers résultats recensés par Google sur un sujet donné jouent un rôle important dans la formation de l'opinion publique¹⁰. Il faut en outre replacer ce fait dans le contexte du passage des médias traditionnels (radiodiffusion, presse) aux médias en ligne, passage provoqué par les besoins en informations de certains groupes d'utilisateurs, mais tout en sachant que les médias de masse traditionnels demeurent importants pour la communication publique¹¹.

Pour revenir aux différences mentionnées plus haut entre médias d'affichage et outils de recherche, certains constatent que la fragmentation de l'espace public a déjà conduit à l'apparition d'un nombre croissant de services adaptés à des intérêts particuliers (et vice versa), et suggèrent qu'elle a été provoquée par les outils de recherche. Tout d'abord, là où les questions reflétant les intérêts d'un individu recoupaient largement celles qui préoccupaient d'autres citoyens, des études empiriques révèlent une modification substantielle, à cet égard, dès lors que les individus s'informent principalement par le biais des outils de recherche.

En outre, il est indéniable que les moteurs de recherche offrent un accès (facile) à des œuvres protégées et à des services spécialisés, tels que la recherche d'images ou de contenus audiovisuels. Dans le cas des images, les fournisseurs de moteurs de recherche jouent un rôle plus actif que lorsqu'ils se contentent de proposer aux utilisateurs des liens vers des œuvres, puisque les moteurs de recherche génèrent des onglets des images. On peut toutefois arguer du fait que ce fonctionnement ne contrevient pas aux lois de protection du droit d'auteur¹². Ceci vaut également, dans une large mesure, pour les contenus audiovisuels.

Les moteurs de recherche comptent parmi les sites les plus visités sur Internet. Les requêtes qui leur sont soumises constituent en outre une excellente base pour le profilage. Par conséquent, les moteurs de recherche suscitent aussi des interrogations en matière de protection des données¹³. Et dès lors que la recherche de portraits photographiques est en jeu, d'autres aspects des droits de la personne peuvent également se trouver menacés.

Dernier point, puisqu'un unique fournisseur de moteur de recherche a acquis une part de marché extrêmement importante dans plusieurs pays, tant dans le domaine des services de recherche que dans celui de la publicité ciblée en ligne, il faut également aborder le problème de la distorsion de la concurrence. Sachant que des sociétés telles que Google étendent leurs activités à d'autres domaines – publicité, par exemple¹⁴ – le transfert de leur pouvoir de marché à d'autres marchés – comme celui de la publicité – prend là aussi une importance croissante.

9) En ce qui concerne le filtrage sur Internet, Voir Deibert/ Palfrey/Rohozinski/Zittrain, *Access Denied: The Practice and Policy of Global Internet Filtering*.

10) Voir Schulz/Held/Laudien, *Search Engines as Gatekeepers of Public Communication Analysis of the German Framework Applicable to Internet Search Engines Including Media Law and Anti-trust Law*, dans : *German Law Journal*, 2005 (vol 6, n° 10), p. 1419-1433, disponible sur : http://www.germanlawjournal.com/pdf/Vol06No10/PDF_Vol_06_No_10_1419-1432_Developments_SchulzHeldLaudien.pdf

11) Voir Ridder/Engel, *Massenkommunikation 2005: Images und Funktionen der Massenmedien im Vergleich*. Dans : *Media Perspektiven*, n° 9/2005, p. 422 et. seq.

12) Voir Wimmers/Schulz, *Wer nutzt? - Zur Abgrenzung zwischen Werknutzer und technischem Vermittler im Urheberrecht*, CR 2008, p. 170 et suivantes ; LG Hamburg (Cour régionale de Hambourg), affaire n° 308 0 449/03, MMR 2006, 697-699 ; LG Erfurt (Cour régionale d'Erfurt), affaire n° 3 0 1108/05, MMR 2007, 393-394 ; commenté par Berberich dans CR 2007, 393-395 et Roggenkamp dans K&R 2007, 328-330 ; OLG Jena (Cour d'appel d'Iéna), jugement du 27 février 2008 - 2 U 319/07. Pour de plus amples informations sur les aspects juridiques des moteurs de recherche, des liens et des cadres, ainsi que sur le droit d'auteur allemand, consulter <http://www.linksandlaw.com> de Stephan Ott.

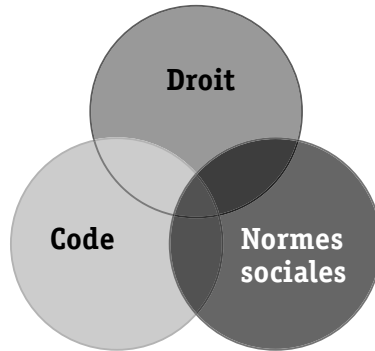
13) Concernant le débat au niveau européen, voir le Groupe de travail « article 29 » sur la protection des données, disponible sur : http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2008/wp148_fr.pdf.

14) Dans ce contexte, notamment, le rachat par Google de DoubleClick est important, disponible sur : http://www.businessweek.com/technology/content/apr2007/tc20070414_675511.htm

3. Approches réglementaires

3.1. Généralités

En partant du principe que l'un des risques mentionnés ci-dessus est réglementé, et que la réglementation actuelle est insuffisante pour traiter les problèmes existants, il faut faire le bon choix en termes de réglementation. Sur un plan très abstrait, le triangle suivant décrit les principes directeurs à prendre en compte dans ce domaine :



Les normes sociales, à l'instar des règles professionnelles dans le journalisme, déterminent certains types de comportements sociaux. Il existe des ensembles de normes qui s'appliquent spécifiquement à Internet (par exemple la netiquette). Des règles explicites peuvent exister pour certains groupes – c'est le cas des codes de conduite ou, de façon moins officielle, des « consignes internes », qui sont beaucoup plus courantes et peuvent être également être efficaces. La réglementation juridique peut se fondre dans ces normes sociales, mais elle peut aussi se trouver érodée ou voir sa mise en œuvre affaiblie, lorsqu'elle empiète sur les normes sociales. Ceci peut se produire en particulier si le législateur omet de tenir compte des normes sociales au moment de l'élaboration des lois.

Sur Internet, notamment, l'architecture logicielle et matérielle impose également des contraintes, que l'on appelle code¹⁵. Mais la réglementation juridique est moins habituée à prendre en compte ce type de code lors de l'élaboration du cadre réglementaire ; ce fait pourrait affaiblir les concepts réglementaires censés permettre à l'avenir de réglementer Internet.

3.2. Corégulation et autorégulation

Les approches spécifiques étudiées dans cet article sont la corégulation et l'autorégulation. Le recours à la corégulation est encouragé par l'UE – explicitement, dans l'article 3 de la Directive « Services de médias audiovisuels » (qui ne concerne pas les moteurs de recherche en ligne), mais aussi d'une manière générale, dans le cadre du programme « Mieux légiférer »¹⁶. Parler de corégulation et d'autorégulation, c'est aborder l'interaction entre la régulation gouvernementale et non gouvernementale, ainsi que la liberté de manœuvre laissée à chaque camp pour définir le concept de régulation. D'un côté, on trouve l'idéal d'une forme traditionnelle de contrôle hiérarchique (*command-and-control-regulation*), entièrement conçu et mis en œuvre par l'État, tandis que de l'autre, on trouve l'autorégulation pure (qui, dans la pratique, est généralement encadrée par la loi, mais, pour la question qui nous occupe, se trouve intégralement sous contrôle non étatique).

L'un des avantages généraux de l'autorégulation réside dans l'expertise des acteurs, dont l'efficacité se déploie pleinement dès lors qu'ils sont en mesure de prendre leurs propres décisions. Cet aspect vient s'ajouter à l'efficacité, au caractère ciblé, à la flexibilité et la large acceptation qui sont autant de caractéristiques de ce type de régulation. Mais l'autorégulation est également critiquée, pour son manque de légitimité démocratique et juridique, son contournement des mécanismes de protection des libertés fondamentales (mise en œuvre d'une censure privée), son injustice et l'incertitude juridique qu'elle entraîne, sans oublier les inquiétudes suscitées quant à la fragmentation du droit.

15) Voir Lessig, CodeV2, disponible sur : <http://codev2.cc/download+remix/>

16) Voir Accord interinstitutionnel « Mieux légiférer » 2003/C 321/01.

La corégulation lie les deux approches mentionnées ci-dessus. Du point de vue du Hans-Bredow-Institut, elle peut être définie comme suit : « La corégulation consiste à associer la régulation gouvernementale et non gouvernementale de façon à ce qu'un système de régulation non gouvernemental établisse des liens avec la réglementation relevant de l'État¹⁷. »

Cette définition précise ensuite les critères que doit remplir un concept réglementaire pour correspondre à la définition de la corégulation.

Ainsi, l'élément non gouvernemental du système de régulation recouvre :

- La création d'organismes, règles et processus spécifiques,
- dans le but de peser sur les décisions,
- prises par les destinataires eux-mêmes.

Concernant les liens entre un système régulateur non gouvernemental et la réglementation de l'État, on peut parler de corégulation dès lors que les critères suivants sont remplis :

- Réalisation d'objectifs de politique publique,
- Liens juridiques,
- Pouvoir discrétionnaire du système régulateur non étatique,
- Utilisation par l'État de ces ressources de régulation¹⁸.

La corégulation requiert un ensemble de conditions. Tout d'abord, une certaine culture de régulation est nécessaire. Ceci est particulièrement important dans le cas des modèles de corégulation qui s'en remettent aux organismes sectoriels, par exemple, pour élaborer des codes ou faire appliquer des règles. Par ailleurs, les mesures d'incitation à la coopération et à la mise en œuvre des règles sont d'une importance cruciale. Il faut que les acteurs du secteur aient une motivation suffisante pour coopérer entre eux. Dans la plupart des cas, c'est la perspective d'éviter une régulation de l'État qui constitue cette incitation. L'État doit par conséquent réduire substantiellement son pouvoir réglementaire ; mais dans le même temps, le secteur doit continuer de penser que l'État mettra en œuvre sa propre régulation si la régulation non gouvernementale échoue.

À côté de ces incitations, un système de corégulation opérationnel doit en outre garantir sa propre mise en œuvre, grâce à des sanctions adaptées et proportionnées. Autre condition pour l'existence d'un système de corégulation, l'État doit conserver des moyens qu'il utilisera pour peser sur le résultat des processus réglementaires qui ne dépendent pas de lui. Ces ressources sont nécessaires, dans la mesure où l'État conserve une responsabilité : garantir un niveau de protection suffisant pour que les objectifs publics soient atteints. Sur ce point, pour les codes et/ou les organismes non étatiques, les procédures de certification ont donné de bons résultats. D'autres moyens envisageables peuvent être la délégation de membres à des organismes non étatiques, les incitations financières ou encore la publicité. Pour finir, une base juridique et un partage des tâches nets sont essentiels en matière de corégulation. Si elle est bien claire, la division des tâches constitue pour le secteur une forte incitation à participer, tandis que si elle est floue, elle peut aboutir à un manque de transparence, augmenter le risque de « piéger » l'organisme de régulation et conduire à des « rideaux de fumée ».

3.4. Exemples de corégulation d'Internet en Allemagne

En Allemagne, les organismes non étatiques ont toujours joué un rôle important en matière de protection des mineurs dans les médias ; ils sont depuis longtemps responsables de la classification des

17) Voir Baldwin/Cave, *Understanding regulation: Theory, Strategy, and Practice* (1999) ; Ayres/ Braithwaite, *Responsive Regulation*, (1992) ; Hans-Bredow-Institut, *Study on co-regulation measures in the media sector*, disponible sur : http://ec.europa.eu/avpolicy/info_centre/library/studies/index_en.htm ; Held/Schulz, *Regulated self-regulation*, disponible sur :

<http://www.hans-bredow-institut.de/publikationen/apapiere/8selfreg.pdf>

18) Voir Schulz/Held, *Together they are Strong?* dans : *Regulation, Awareness and Harmful Media Content in the Digital Age*, Ulla Carlsson (Ed.) (2006), p. 49 et seq. (pages 51 et 52), disponible sur : <http://unesdoc.unesco.org/images/0014/001469/146955E.pdf>

contenus en fonction de l'âge. La *Jugendschutzgesetz* (loi fédérale sur la protection des mineurs – JuSchG) établit une distinction entre différents niveaux de contenus et prévoit que ceux qui sont préjudiciables pour la jeunesse (*jugendgefährdend*) fassent l'objet d'une classification par la *Bundesprüfstelle für jugendgefährdende Medien* (Autorité fédérale de surveillance des médias préjudiciables aux mineurs – BPjM). La mission de la BPjM consiste à protéger les enfants et les jeunes en Allemagne contre tout média susceptible de présenter des contenus préjudiciables ou dangereux. Elle surveille notamment les médias suivants : vidéos, DVD, jeux vidéo, disques et CD, médias papier et sites Internet. Ces supports sont jugés préjudiciables ou dangereux pour les mineurs dès lors qu'ils tendent à entraver le processus de développement d'une personnalité responsable et autonome. De manière générale, ceci s'applique aux contenus indécents, extrêmement violents, criminogènes, à caractère antisémite ou raciste. Si un support est jugé dangereux pour les mineurs (*jugendgefährdend*), son titre est placé sur la « liste des médias dangereux pour la jeunesse » (*Liste jugendgefährdender Medien, Index*). Dès lors, les distributeurs de ces médias ne sont plus autorisés à vendre, louer ou même présenter en public ces contenus, ou même à les radiodiffuser. Le contenu préjudiciable pour les mineurs ne doit pas être présenté dans des lieux auxquels les enfants ont accès ; les mineurs ne doivent pas pouvoir se le procurer. Dans le domaine des films et des jeux vidéos, un contenu qui n'est pas nuisible aux enfants, mais est susceptible d'entraver leur développement (*entwicklungsbeeinträchtigt*), fait l'objet d'une classification par l'*Oberste Landesjugendbehörde* (haute autorité du Land pour la jeunesse).

Cette classification par âge s'effectue dans le cadre d'une coopération avec des organismes non étatiques : le *Freiwillige Selbstkontrolle Filmwirtschaft* (organisme d'autocontrôle volontaire du secteur cinématographique – FSK) et l'*Unterhaltungssoftware Selbstkontrolle* (organisme d'autocontrôle des logiciels de loisir – USK). Les organismes gouvernementaux et non gouvernementaux sont étroitement liés¹⁹.

Suivant cette tradition, les moteurs de recherche ont mis sur pied un système d'autocontrôle entièrement fondé sur le volontariat. Le succès croissant et l'enrichissement permanent des moteurs de recherche leur confèrent une portée accrue en matière de protection de la jeunesse dans les médias. Conscients de ces faits, de nombreux grands fournisseurs de moteurs de recherches se sont rencontrés sous l'égide d'un autre organisme de régulation, le *Freiwillige Selbstkontrolle Multimedia-Diensteanbieter* (organisme d'autocontrôle volontaire des fournisseurs de services multimédia – FSM) et ont créé en février 2005 *Selbstkontrolle Suchmaschinen* (organisme d'autocontrôle des fournisseurs de moteurs de recherche)²⁰, une initiative qui s'appuie sur un code de conduite commun. Son objectif est d'améliorer la protection des consommateurs, ainsi que la protection des enfants et des jeunes qui utilisent les moteurs de recherche en Allemagne. Aux termes de ce code, les fournisseurs de moteurs de recherche acceptent d'expliquer à leurs utilisateurs le mode de fonctionnement de l'outil de recherche. Ils s'accordent également pour structurer de façon transparente les pages de résultats de recherche. En outre, les signataires de ce code doivent s'efforcer, dans la mesure de leurs moyens, de permettre la prise de certaines précautions techniques permettant de protéger les enfants et les jeunes contre des contenus qui leur sont préjudiciables. De plus, les fournisseurs de moteurs de recherche intègrent dans leur outil de recherche le « module BPjM » fourni par la BPjM. La diffusion des contenus pornographiques et, généralement, préjudiciables, via Internet, constitue une infraction pénale en droit pénal allemand. Les contenus pornographiques sur Internet ne sont légaux que dans la mesure où des mesures techniques empêchent l'accès par les mineurs (système de vérification de l'âge ou « *Adult check system* », AVS). Les contenus préjudiciables aux mineurs (*jugendgefährdend*) sont placés dans l'« index » mentionné plus haut. Le « module BPjM » est une liste de services Internet étrangers qui contiennent des contenus préjudiciables, conservée par la BPjM et communiquée au FSM. Ce dernier peut utiliser cette liste pour mettre en place des mesures de sécurité (filtrage), et les fournisseurs, comme on l'a signalé plus haut, ont accepté de retirer et/ou de ne pas afficher toute URL placée sur la liste par la BPjM, dans la mesure où les fournisseurs de moteurs de recherche ont effectivement accès à cette URL, et où cette action entraîne des dépenses commerciales raisonnables. En cas de violation de ce code de conduite, le contrevenant s'expose à des sanctions. Les moteurs de recherches ont accepté de ne pas afficher ces résultats sur leurs services en nom de domaine « .de ».

19) Voir Schulz/Held, *Together they are Strong?* dans : *Regulation, Awareness and Harmful Media Content in the Digital Age*, Ulla Carlsson (Ed.) (2006), p. 49 et seq. (57 et seq.), disponible sur : <http://unesdoc.unesco.org/images/0014/001469/146955E.pdf> ; sur la question de l'autocensure, Held/Schulz, *Der Index auf dem Index? Selbstzensur und Zensur bei Suchmaschinen*, dans Machill/Beiler, *Die Macht der Suchmaschinen – The Power of Search Engines* (2007).

20) Voir http://www.fsm.de/en/Search_Engines

4. Conclusion : la régulation dans le domaine des moteurs de recherche

Il paraît utile de revenir à la liste de risques générés par les moteurs de recherche et d'appliquer les critères déterminant dans quels cas la corégulation et l'autorégulation seraient opportunes ; mais il reste quelques domaines où un examen plus attentif semble crucial.

L'exemple allemand montre ainsi que la corégulation et l'autorégulation pourraient être efficaces dans le domaine de la protection des mineurs. Ceci vaut aussi pour les problèmes de discrimination envers certains contenus et de tromperie des consommateurs. Là encore, l'autorégulation volontaire des fournisseurs de moteurs de recherche en Allemagne permet d'ores et déjà de répondre à certaines questions concernant la transparence du processus de sélection (mais pas l'algorithme en tant que tel). Quant au risque de voir les moteurs de recherche jouer un rôle dans l'exploitation d'œuvres (audiovisuelles) protégées ou de données personnelles, il semble qu'il y ait également une certaine marge de manœuvre pour mettre en œuvre une corégulation.

Toutefois, dès lors qu'il est question d'objectifs de politiques publiques – tels que le contrôle de l'influence sur le façonnement de l'opinion publique, ou la fragmentation de l'espace public, qui pourrait se trouver aggravée par les moteurs de recherche – les fournisseurs de moteurs de recherche ne sont nullement incités à coopérer. En outre, la distorsion de la concurrence et le transfert des pouvoirs de marché sont bien évidemment des domaines où l'on ne s'attend pas à ce que les prestataires de services proposent spontanément leur coopération. A cet égard, si une quelconque régulation s'avère nécessaire, c'est celle de l'Etat, traditionnelle, qui semblerait s'imposer.

Lors de l'étude des différentes approches en matière de régulation, il faut toutefois garder à l'esprit le fait que les fournisseurs de moteurs de recherche ne sont pas pleinement conscients du rôle qu'ils jouent dans les communications publiques. Les fournisseurs de moteurs de recherche sur Internet s'intéressent essentiellement à l'élaboration de la technologie de recherche. Dans la conduite de leurs activités, ils estiment qu'ils se contentent de créer des outils pour les utilisateurs et ne jouent donc qu'un rôle limité dans la structuration de la communication publique. Le concept d'« intermédiaires en information », assumant des fonctions particulières en matière de communication publique, est relativement nouveau, y compris dans les débats universitaires. On peut supposer qu'il en est de même pour les professionnels travaillant dans ce domaine. Quoi qu'il en soit, les dernières évolutions doivent être jugées positives, parce qu'il est encore possible de mettre en place les déontologies professionnelles correspondantes et que la régulation juridique peut également jouer un rôle vital dans leur instauration.

En outre, tous les fournisseurs de services sont intéressés par l'externalisation de la responsabilité, dès lors qu'il s'agit de faire des choix en matière de contenus. En conséquence – pour reprendre le même exemple – la liste de sites comprenant des contenus préjudiciables fournie par la BPJM allemande a été accueillie chaleureusement par le secteur, tout simplement parce qu'elle évite aux acteurs du secteur d'avoir à décider quels sont les sites qu'ils doivent bloquer.

Pour finir, les fournisseurs exerçant à l'échelle internationale ne disposent que d'un champ limité pour réagir aux initiatives nationales de régulation. L'autorégulation et la corégulation sont en mesure de fournir des outils flexibles pour une coopération internationale, tout en permettant simultanément l'élaboration de solutions mondiales et en inscrivant ce processus dans l'évolution réglementaire propre à chaque pays. Par conséquent, il demeure essentiel que le code de conduite puisse être adapté aux conditions locales (les moteurs de recherche permettant des variations d'algorithmes).

La recherche de contenus audiovisuels

Participants à l'atelier du 12 avril 2008 à Amsterdam

Christina	ANGELOPOULOS	Institut du droit de l'information (IViR), NL (rapporteur)
Monica	ARIÑO	Ofcom, UK
Pierre-Jean	BENGHOZI	CRC de l'Ecole polytechnique, FR
Machiel	BOLHUIS	European Policy Council, Google, UE
Francisco	CABRERA	Observatoire européen de l'audiovisuel, CoE
Ramón	COMPañÓ	IPTS, ES
Anja	DOBBELSTEEN	Institut du droit de l'information (IViR), NL (logistique)
Nico van	EIJK	Institut du droit de l'information (IViR), NL (modérateur)
Urs	GASSER	Université de St Gallen, CH
Janet	GRECO	Broadcastprojects, UK
Natali	HELBERGER	Institut du droit de l'information (IViR), NL
Joris van	HOBOKEN	Institut du droit de l'information (IViR), NL (rapporteur)
Erik	JOSEFSSON	Electronic Frontier Foundation, US
Eddan	KATZ	Electronic Frontier Foundation, US
Miha	KRISELJ	APEK, SI
Tarlach	MCGONAGLE	Institut du droit de l'information (IViR), NL
Milton	MUELLER	Université Syracuse, USA, Université Technique, NL
Susanne	NIKOLTCHEV	Observatoire européen de l'audiovisuel, CoE
Michael	ROTERT	EuroISPA, BE
Thomas	ROUKENS	Telenet, BE
Alexander	SCHEUER	Institut du droit européen des médias (EMR), DE
Wolfgang	SCHULZ	Hans-Bredow-Institut, DE

David	STEVENS	ICRI, Université Leuven, BE
Ralph	TRAPHOENER	Empolis, DE
Peggy	VALCKE	ICRI, Université Leuven, BE
Erik	VALGAEREN	Stibbe, BE
Adam	WATSON-BROWN	DG Infosoc, Commission européenne, UE
Michael	ZIMMER	Université de Wisconsin-Milwaukee, US

Qu'êtes-vous en droit d'attendre d'IRIS Spécial ?

Les publications de la série IRIS Spécial de l'Observatoire européen de l'audiovisuel font le point complet d'une situation, assorti d'analyses approfondies. Concernant le choix des thèmes analysés, nous vous proposons des sujets actuels du droit des médias, développés sous un angle juridique. IRIS Spécial vous présente trois volets qui, selon le thème, peuvent être mis en corrélation :

1. Un exposé précis des bases légales aux fins de comparaison du droit en vigueur dans les différents pays. Ainsi le numéro d'IRIS Spécial « Les obligations des radiodiffuseurs d'investir dans la production cinématographique » procède-t-il à un état des lieux dans 34 pays européens.
2. L'analyse de problématiques actuelles, dont les développements, les tendances juridiques et des ébauches de solutions possibles. Dans cet esprit, le numéro d'IRIS Spécial « Services de médias audiovisuels sans frontières » vous propose une approche axée sur l'avenir, qui conservera son actualité bien après l'adoption de la directive européenne.
3. Une description du cadre réglementaire européen ou international transposé dans le droit national. IRIS Spécial « Avoir ou ne pas avoir- Les règles du must-carry » explique le modèle européen et le compare avec l'approche américaine.

Quelles sont nos compétences ?

Chaque numéro d'IRIS Spécial est conçu par le Département Informations juridiques de l'Observatoire européen de l'audiovisuel, en collaboration avec ses organisations partenaires et un vaste réseau de spécialistes du droit des médias. Les thèmes sont préparés par les participants à des séminaires-ateliers ou des auteurs spécialisés dont nous sollicitons la collaboration. Les participants aux ateliers et les auteurs sont choisis de manière à obtenir un panel représentatif en termes d'horizons nationaux et culturels, de professions et de spécialités.

La valeur ajoutée d'IRIS Spécial

Les publications d'IRIS Spécial abordent les thèmes juridiques de telle sorte que les dossiers soient accessibles à des non juristes. Chaque numéro allie la rigueur scientifique à un fort ancrage dans la réalité.

IRIS Spécial se concentre en premier lieu sur l'Europe, mais élargit l'horizon géographique lorsque le thème s'y prête.

Qui devrait lire IRIS Spécial ?

IRIS Spécial est incontournable si vous occupez un poste de responsable dans l'industrie des médias, notamment dans les secteurs du cinéma, de la radiodiffusion et des nouveaux médias, si vous êtes chargé de la politique médias et culture, spécialiste des médias dans un cabinet d'avocats, dans une instance de régulation ou dans le milieu universitaire.

Déjà parus dans la série IRIS Spécial :

■ La responsabilité éditoriale

ISBN 978-92-871-6475-9 - 52 pages - © 2008 - EUR 75
<http://www.obs.coe.int/editresp>

■ Les aspects juridiques de la vidéo à la demande

ISBN 978-92-871-6390-5 - 97 Pages - © 2008 - EUR 89
<http://www.obs.coe.int/vod-legal>

■ La culture de service public de radiodiffusion

ISBN 978-92-871-6187-1 - 194 Pages - © 2007 - EUR 62,50
<http://www.obs.coe.int/psb>

■ Services de médias audiovisuels sans frontières

ISBN 978-92-871-6114-7 - 78 pages - © 2006 - EUR 58,50
<http://www.obs.coe.int/services-medias-av>

■ Les obligations des radiodiffuseurs d'investir dans la production cinématographique

ISBN 978-92-871-5969-4 - 133 pages - © 2006 - EUR 62,50
<http://www.obs.coe.int/tv-invest-cine>

■ Avoir ou ne pas avoir - Les règles du must-carry

ISBN 978-92-871-5896-3 - 54 pages - © 2005 - EUR 58,50
<http://www.obs.coe.int/must-carry>

■ Le futur de la distribution des services audiovisuels

ISBN 978-92-871-5706-5 - 54 pages - © 2005 - EUR 35
<http://www.obs.coe.int/distribution-tv>

■ Débat politique et rôle des médias

ISBN 978-92-871-5674-7 - 131 pages - © 2005 - EUR 44
<http://www.obs.coe.int/debat-politique>

■ La régulation de l'accès à la télévision numérique

ISBN 978-92-871-5400-2 - 139 Pages + 28 pages Glossaire
© 2004 - EUR 44
<http://www.obs.coe.int/acces-tv-numerique>

■ La corégulation des médias en Europe

ISBN 978-92-871-5140-7 - 156 pages - © 2003 - EUR 27
<http://www.obs.coe.int/coregulation>

■ La compétence juridique en matière de radiodiffusion en Europe

ISBN 978-92-871-4853-7 - 64 pages - © 2002 - EUR 27
<http://www.obs.coe.int/competence-juridique>

■ La télévision et la concentration des médias

ISBN 978-92-871-4594-9 - 97 pages - © 2001 - EUR 27
<http://www.obs.coe.int/concentration-media>