Google Print : changement de support ou changement de paradigme ?

http://print.google.com

Par Olivier Vilaça

Merci à Frederick Fenter1 pour sa relecture précieuse et compétente de cet article (ce qui n'engage en rien ses propres idées !)

Google Print, un des multiples projets lancés par l'entreprise Google, consiste à scanner et à rendre accessible sur Internet l'ensemble de l'information contenue actuellement dans les livres. Cette initiative, qui suscite de nombreuses oppositions, mais aussi des projets concurrents, est susceptible d'avoir un impact important sur les pratiques de recherche et sur l'industrie du livre. Cet article a pour ambition de présenter le point de vue d'un chercheur qui essaie d'entrevoir ce que pourrait être le monde après Google Print.

Google Print : qu'est-ce que c'est ?

Pour bien comprendre ce qu'est Google Print, il suffit de se reporter à la présentation qu'en donne l'entreprise californienne sur la page d'accueil du projet. En quelques mots, on saisit rapidement l'intérêt et les enjeux de ce projet titanesque :

« *Google's mission is to organize the world's information, but much of that information isn't yet online. Google Print aims to get it there by putting book content where you can find it most easily – right in your Google search results.* »

Tout est dit. Le livre est vu comme un support particulier pour stocker l'information. L'objectif de Google est de devenir une interface incontournable pour avoir accès à cette information.

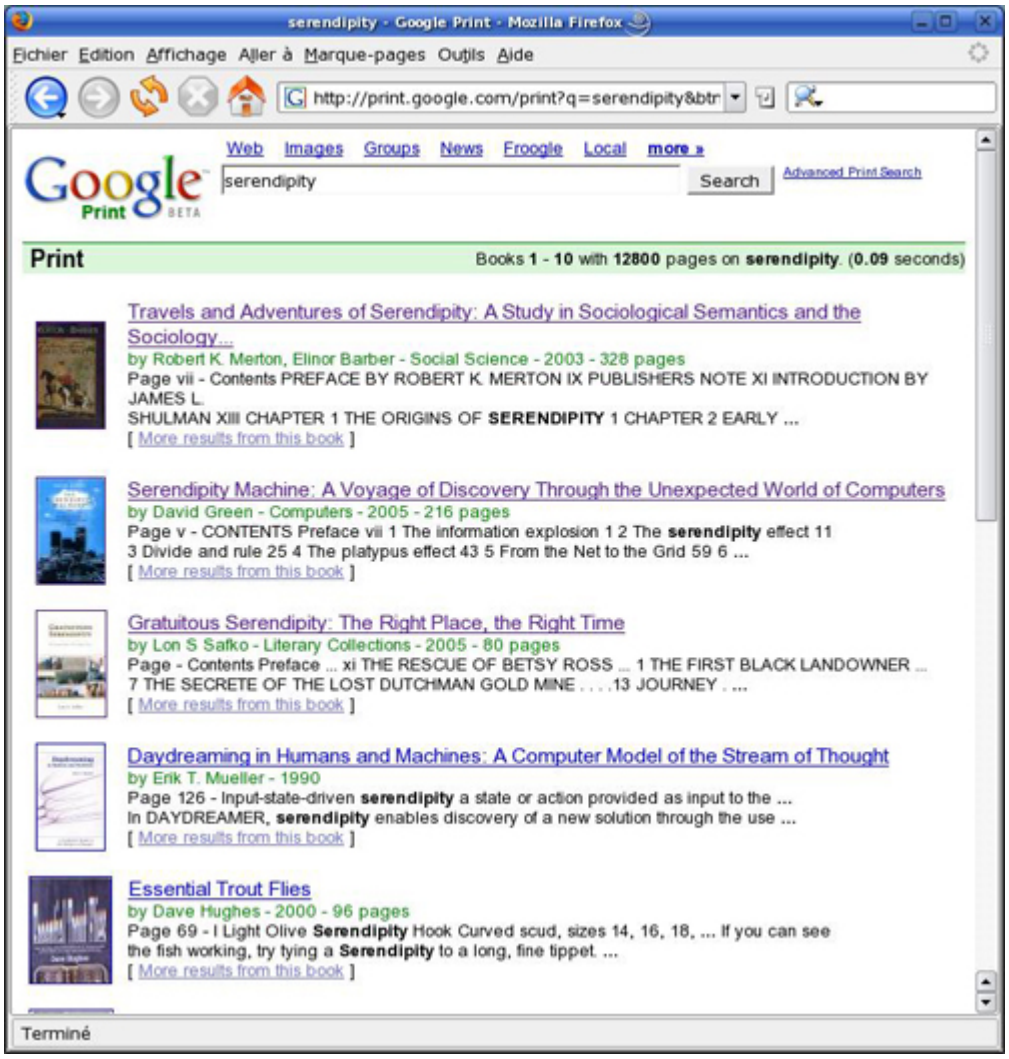
Les sources de Google Print sont à la fois des bibliothèques (pour tous les livres qui ont déjà été publiés) et des maisons d'éditions (pour tous les livres à venir).

L'entreprise a passé un contrat avec cinq bibliothèques : la bibliothèque publique de New York et les bibliothèques des universités d'Oxford, Harvard, Stanford, et de l'Université du Michigan pour numériser leur fond et le rendre accessible via son moteur de recherche. Dans ce cas, ce sont les bibliothèques qui sélectionnent les livres qui doivent faire l'objet de ce traitement. D'autre part, Google a mis en place un programme réservé aux éditeurs qui souhaiteraient participer au projet, en choisissant eux-mêmes les livres à inclure dans la base de données de Google Print.

Comment fonctionne Google Print ?

Le fonctionnement de Google Print est élémentaire puisqu'il est identique à celui du moteur de recherche de Google. La seule différence, est qu'au lieu de chercher des occurrences sur le World Wide Web ou dans les forums de discussions, Google va scruter une base de données uniquement composée de livres. Contrairement à une base de données bibliographique (tel qu'Opale plus, la base de données de la Bibliothèque Nationale Française [Bnf], par exemple), la recherche ne se fait pas en fonction de champs spécifiques et limités (mot du titre, nom de l'auteur, mot du résumé ou mot-clé) mais dans l'intégralité du texte du livre. Ce dernier point est particulièrement important pour le chercheur, car il est susceptible de révolutionner ses pratiques de recherche comme on le verra dans la suite de cet article.

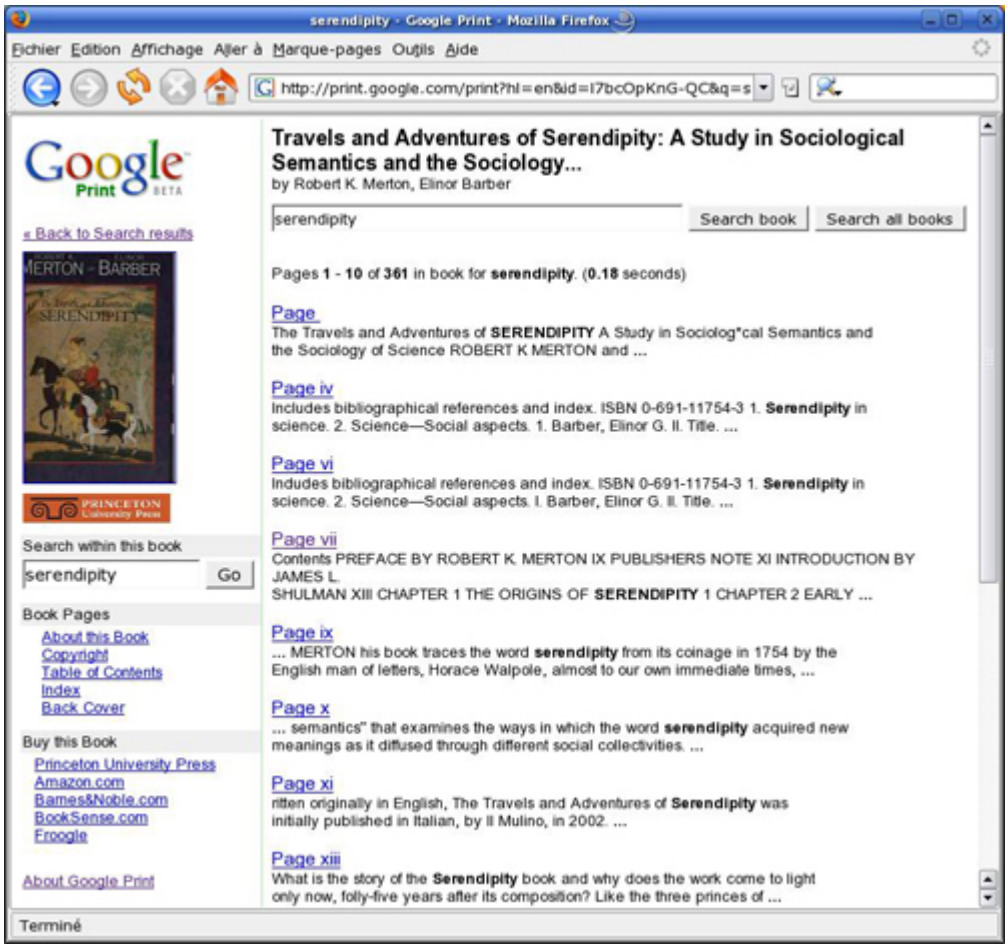
Une fois la recherche effectuée, Google Print affiche la liste de tous les livres dont au moins une page contient l'objet de la requête. Les options de recherche sont les mêmes que sur Google, c'est-à-dire que l'on peut rechercher un ou plusieurs mots, ou encore une phrase entière (pratique pour retrouver une citation). Il est également possible d'exclure des mots ou des phrases pour affiner les résultats. Mais Google Print ne s'arrête pas là. Une fois qu'un livre est sélectionné, une nouvelle option de recherche est proposée qui permet de naviguer dans ce livre, permettant ainsi de l'explorer de fond en comble.



Prenons maintenant un exemple concret. Imaginons que je cherche des ouvrages traitant de la sérendipité. Il me suffit pour cela de me rendre sur http://print.google.com et d'entrer le mot « serendipity » dans le moteur de recherche (puisque la majeure partie des livres présents sur Google Print sont en anglais).

Quelques dixièmes de secondes plus tard, j'obtiens une page m'affichant les 10 premiers résultats d'une liste de 12800 livres contenant l'occurrence recherchée.

Si je clique sur l'option « more results from this book » du premier résultat disponible, *Travels and Adventures of Serendipity*, j'obtiens alors la liste de toutes les pages de ce livre contenant le mot « serendipity ».



Je peux alors éventuellement naviguer dans ce livre en utilisant l'option « search within this book », afin d'explorer plus en détail son contenu et voir s'il a un intérêt pour moi. On peut alors d'ores et déjà imaginer plusieurs usages différents à Google Print.

À quoi sert Google Print ? (1)

Le premier usage possible consisterait à utiliser Google Print pour lire des livres, soit directement à l'écran, soit en les imprimant. Autant le dire tout de suite, cet usage n'a que très peu d'intérêt car, bien que Google ait fait le choix de reprendre la mise en page des livres tel quel à l'écran, lire un livre sur un ordinateur n'a pas de véritable avantage par rapport au fait de le lire sur un support papier. Le livre, contrairement à l'ordinateur, est facile à transporter, il permet également une lecture d'autant plus confortable que son contenu est long. On peut lire un livre dans n'importe lieu, dans n'importe quelle situation et n'importe quelle position, ce qui est très loin d'être évident avec un ordinateur,

fusse-t-il portable. La solution qui consisterait à imprimer le livre à partir de Google n'est pas vraiment satisfaisante. Elle est coûteuse et génère un bloc de feuilles peu pratique à manier.

D'autre part, il faut préciser que tous les livres présents sur Google Print ne sont pas accessibles dans leur intégralité. En effet, les conditions d'accès aux documents varient en fonction de leur statut et des accords que Google a signés avec les maisons d'édition. Il est ainsi possible de distinguer trois types de documents :

- Ceux qui sont tombés dans le domaine public. Leur contenu est alors accessible dans son intégralité.

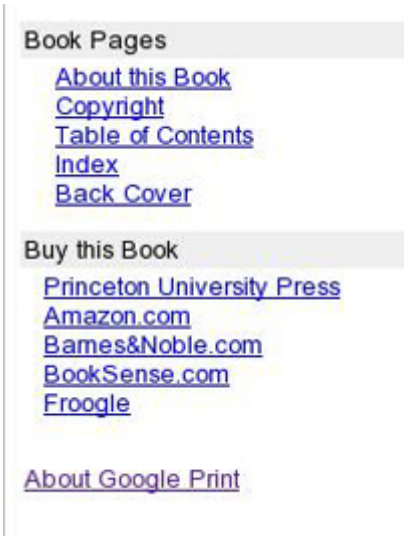
- Ceux dont les droits de reproduction sont encore réservés mais qui ont fait l'objet d'un accord entre la société Google et la maison d'édition qui les a publiés. Dans ce cas, l'internaute peut visualiser deux pages en plus de celle qui contient les résultats de sa requête, le nombre de pages totales visualisées étant limité.

- Ceux qui sont encore sous copyright et qui n'ont pas fait l'objet d'un accord entre Google et la maison d'édition qui les a publiés. L'accès à ce type de document est limité à quelques lignes.

Google Print n'est donc pas une gigantesque bibliothèque gratuite, même si l'heureux possesseur d'un compte Google (pour accéder au service de messagerie Gmail par exemple) dispose d'un accès étendu au contenu des livres. Il est toutefois possible de tirer deux leçons de ce qui vient d'être dit. Tout d'abord, le livre est à la fois un objet et un contenu. C'est un objet auquel de nombreuses personnes sont attachées. Le livre se sent, se touche, est souvent beau et s'expose dans une bibliothèque ou sur la table du salon. C'est aussi un contenu, c'est-à-dire un ensemble d'informations couchées par écrit sur un support particulier qui s'est révélé être le plus efficace jusqu'à aujourd'hui pour

permettre l'accès de cette information au plus grand nombre et sur le long terme. Le codex, c'est-à-dire le livre est, par exemple, plus pratique que le volumen (ou rouleau) utilisé autrefois. Ensuite, on peut se demander si la forme du livre, qui a un début, un milieu et une fin, ne détermine pas son contenu. La question que cette remarque soulève est donc la suivante : notre réflexion de scientifique serait-elle aussi linéaire si elle n'était pas destinée à un moment ou à un autre à être publiée, ou en tout cas à être couchée par écrit ?

À quoi sert Google Print ? (2)



Le deuxième usage possible, bien plus pratique celui-ci, consisterait à utiliser Google Print comme un catalogue géant, un formidable index qui, parce qu'il permet l'accès au contenu des livres, permet d'identifier plus rapidement et avec plus de pertinence les ouvrages que l'on souhaiterait par la suite acquérir sous leur forme papier. Google répète ainsi à qui veut l'entendre que son objectif n'est pas de détruire l'industrie du livre puisque, d'ailleurs, des liens vers des sites de librairies en ligne sont systématiquement proposés dans ses pages de résultats.
Comme on l'a vu précédemment, l'avantage d'avoir accès à l'intégralité du texte par l'intermédiaire d'un moteur de recherche permet d'identifier avec précision un ouvrage, sans avoir à deviner son intérêt à partir de quelques mots du titre ou du résumé. Cet usage de Google Print est aujourd'hui le plus évident, ce qui est normal étant donné que l'industrie du livre est encore prépondérante. Cependant, il y a un décalage assez frappant entre ce simple rôle de faire-valoir

commercial et l'ambition de la mission que Google dit s'être fixée, à savoir organiser l'information mondiale. Soit on pense que les dirigeants de Google ont une courte vue et qu'ils veulent simplement trouver tous les moyens possibles et imaginables pour générer des recettes publicitaires, ce que permettrait Google Print comme n'importe quel autre projet, soit on pense qu'ils ont une vision plus ambitieuse de l'avenir où l'information — c'est principalement vrai pour l'information scientifique — ne serait plus contrôlée par des maisons d'éditions, mais par des moteurs de recherche. Pure fiction ? Je pense que c'est une hypothèse sur laquelle il est intéressant, voire même important de se pencher dès maintenant.

À quoi sert Google Print ? (3)

Le troisième usage possible, qui reste largement à explorer, consisterait à utiliser Google Print pour effectuer une des tâches les plus importantes d'un travail de recherche : la recherche bibliographique. Pour bien comprendre cet usage, et l'enjeu qu'il suscite, il est tout d'abord nécessaire de rappeler en quoi consiste la recherche bibliographique.

Comme le rappelle justement Google, l'essentiel de l'information mondiale disponible aujourd'hui se trouve dans les livres. Le chercheur est donc logiquement amené à explorer ces sources d'informations qui sont nécessaires à toutes les étapes de sa recherche (émettre des hypothèses, les valider, proposer une théorie ou un modèle). Les livres peuvent même constituer le « terrain » du chercheur, comme c'est souvent le cas en histoire. Dans tous les cas, il est assez facile de découper les différentes étapes nécessaires à l'accomplissement de cette tâche.

L'ordre des deux premières étapes peut être inversé. Quoi qu'il en soit, le chercheur doit, à un moment ou à un autre, se rendre dans une bibliothèque. Cette étape n'est pas anecdotique, puisqu'elle entraîne des coûts non négligeables (de transport et d'inscription) et qu'elle est chronophage (temps de déplacement, temps d'attente du livre lorsqu'il se trouve en magasin). Il doit également identifier les livres qui sont susceptibles de l'intéresser. Il peut alors bénéficier des conseils de ses pairs ou identifier une référence dans un livre qu'il

a déjà lu. Mais, dans la plupart des cas, il est amené à consulter une base de données bibliographique ou un index et deviner, à partir d'un nombre de mots très réduits (mots du titre ou, dans le meilleur cas, mots du résumé), quels livres ont un intérêt pour lui. Cette dernière étape demande un savoir-faire particulièrement technique.

Une fois le tri réalisé, vient la troisième phase qui est encore plus technique, aléatoire et chronophage, et qui consiste à explorer le contenu des livres identifiés (lorsqu'ils sont disponibles bien sûr). Les livres que le chercheur lit en entier, c'est-à-dire de façon linéaire, sont généralement relativement rares. Ils constituent alors les ouvrages de référence sur lesquels le chercheur va se baser pour construire son propre raisonnement. Mais, la plupart des lectures du chercheur seront plutôt partielles, et se feront selon des techniques de lecture en diagonale, qui prennent malgré tout du temps et qui laissent pas mal de déchets. Ici, le chercheur cherche avant tout à disposer du meilleur rapport entre le temps passé à lire et le nombre d'informations prélevées, même si, parfois, cela peut se faire au détriment d'une information importante.

C'est à ce stade de la réflexion qu'intervient Google Print. L'innovation majeure de ce projet ne réside pas seulement dans le fait que les livres soient accessibles en ligne — ce qui limite déjà les contraintes d'accessibilité à la bibliothèque en tant que lieu — mais aussi dans le fait que l'ensemble du contenu des livres peut faire l'objet d'une requête par l'intermédiaire d'un moteur de recherche. Cela a au moins trois conséquences sur les pratiques de recherche. Tout d'abord, le chercheur n'a plus besoin de se déplacer puisque l'ensemble du contenu est disponible en ligne (à condition d'avoir un compte Google). Ensuite, le chercheur peut identifier avec nettement plus de précision et moins de déchets les livres qui sont pertinents et importants pour sa recherche. Enfin, l'option de recherche à l'intérieur d'un livre spécifique permet une lecture en diagonale beaucoup plus rapide et efficace que sur un support papier.

C'est certainement ce dernier point qui est le plus important puisque la lecture du livre se fait par l'intermédiaire de requêtes précises, sans tenir aucunement compte de l'organisation linéaire du livre. On ne tourne plus les pages, on effectue des requêtes sur un moteur de recherche. Cette opération modifie les

pratiques de consultation de l'information. Reste à voir si, et de quelle façon, elle est susceptible de modifier la production et l'organisation de cette information.

Les opposants et les concurrents à Google Print (qui sont souvent les deux à la fois).

Les enjeux liés à Google Print sont encore plus visibles et faciles à comprendre lorsque l'on passe en revue les positions des opposants à ce projet — qui peuvent varier assez sensiblement en fonction des intérêts des uns et des autres — et les initiatives concurrentes.

L'opposition la plus tranchée est celle d'un certain nombre d'auteurs qui ont décidé de poursuivre Google en justice pour violation des droits d'auteur. La plainte a en fait été déposée en septembre par une personne morale (la Guilde des Auteurs qui compte environ 8000 membres) et trois personnes physiques (les écrivains Herbert Mitgang, Betty Miles et Daniel Hoffman).

Ils viennent d'être rejoints par l'Association des Editeurs Américains, à laquelle appartiennent McGraw-Hill, Pearson Education, Penguin, Simon & Schuster et John Wiley & Sons, toujours pour les mêmes questions de droits d'auteur. Mais, les maisons d'éditions ne se contentent pas d'essayer de stopper le projet. Des éditeurs allemands se sont ainsi regroupés pour lancer un projet concurrent qui répond au nom de Volltextsuche Online.

Outre les auteurs et les éditeurs, on retrouve également une certaine opposition de la part des bibliothèques, ou, tout au moins, de certaines bibliothèques. Jean-Noël Jeanneney, le président de la Bnf, a ainsi décidé de mener la fronde après avoir été, dans un premier temps, favorable à l'initiative de Google. Ses motivations sont diverses, mais il semble principalement mettre en avant trois arguments qui le conduisent à proposer le lancement d'un projet européen de numérisation des livres. Le premier argument est le refus d'un monopole de Google sur le contrôle de l'accès à l'information. Le deuxième concerne la préservation de la diversité culturelle, Google Print renvoyant principalement à des informations produites par le monde anglo-saxon. S'il est vrai que les bibliothèques partenaires du projet Google Print sont uniquement anglo-

saxonnes (ce qui ne veut d'ailleurs pas dire que leur fonds est uniquement anglophone), le volet qui concerne les éditeurs est, lui, largement ouvert à des maisons d'éditions non anglo-saxonnes. Google a ainsi récemment annoncé l'extension de son programme à huit pays européen (Italie, Allemagne, Hollande, Autriche, Suisse, Belgique, Espagne et France). Le troisième et dernier argument concerne les risques de dérives marchandes liées au produit, notamment la présence de publicités sur les pages de résultats, et le risque de voir les entreprises multinationales influer sur le classement des résultats. Jean-Noël Jeanneney n'est donc en rien opposé à l'idée de numériser et de mettre en ligne le contenu de livres — ce que la Bnf a déjà commencé à faire avec Gallica, un projet beaucoup moins novateur, ambitieux et performant que Google Print. Ce qui semble le perturber, c'est que ce soit Google qui le fasse. Il a donc proposé de lancer une initiative concurrente à l'échelle européenne. L'idée de cette « très grande bibliothèque » virtuelle à même été reprise par Jacques Chirac, le Président français, ramenant ainsi, et une fois encore, une question fondamentale de société à une simple compétition géopolitique.

Enfin, d'autres projets concurrents voient le jour de la part d'autres grands acteurs du monde de l'informatique en général et d'Internet en particulier. Amazon, le célèbre portail marchand qui a été aussi la première librairie en ligne, a déjà lancé un projet de numérisation similaire appelé « Search inside ». Celui-ci permet aux possesseurs d'un compte Amazon ayant déjà effectué des achats en ligne d'accéder aux mêmes fonctions que Google Print. Cependant, le fond proposé par Amazon est sensiblement différent de celui de Google Print puisqu'il ne contient que des ouvrages récents.

Une autre initiative, plus ambitieuse, vient de voir le jour avec la création d'un consortium appelé « Open Content Alliance » (Oca), réunissant des entreprises (Yahoo !, Adobe, Hewlett-Packard, Msn, etc.), des universités (Columbia, Toronto, John Hopkins, etc.), des bibliothèques, des archives (The Internet Archive, les archives nationales de Grande-Bretagne), des éditeurs (O'Reilly), etc. L'objectif de l'Oca est pratiquement le même que celui de Google Print. Cependant, les membres du consortium ont pris les précautions nécessaires pour éviter les mêmes déboires que Google sur la question du copyright. La Faq du site de l'Oca précise ainsi que la mise ne ligne du contenu d'un livre ne se

fera pas sans l'autorisation du détenteur des droits de reproduction, c'est-à-dire, dans la plupart des cas, les maisons d'éditions, et en aucun cas les bibliothèques.

Cependant, il n'est pas certain que cette condition suffise à désamorcer d'éventuelles poursuites contre l'Oca. Si l'éditeur, en tant que propriétaire du copyright, peut autoriser l'accès gratuit à l'intégralité du contenu d'un livre, il n'est pas du tout sûr que cela plaise à son auteur, qui reste malgré tout le propriétaire moral de l'ouvrage. De ce point de vue, il n'est pas anodin que les premières poursuites judiciaires aient été engagées par des auteurs et non par des éditeurs. Mais cette condition pourrait aller jusqu'à signifier l'échec programmé de ce projet tant le nombre de demandes d'autorisations est théoriquement élevé.

D'un point de vue purement pragmatique, voire même utilitariste, l'intérêt du chercheur est d'avoir accès au maximum d'informations possible, et de les consulter le plus rapidement et le plus efficacement possible (et pour un coût restreint puisque cela entre en compte dans les conditions d'accès à l'information). Il aura donc intérêt à ce que le projet qui lui amène tout cela à la fois soit celui qui sortira gagnant de cette lutte engagée pour le contrôle de l'information.

Ce que l'on voit avec ces différentes oppositions et initiatives concurrentes, c'est que le passage d'un type de support (papier) à un autre (numérique) n'est pas une simple opération technique. La numérisation et la mise en ligne du contenu des livres impliquent l'émergence de nouveaux acteurs dont le métier n'est pas l'édition. Il implique peut-être également, et à terme, la disparition des acteurs traditionnels liés à l'industrie du livre dont le modèle économique et le principe d'existence sont radicalement remis en question par Google Print en particulier, et Internet en général.

Si aujourd'hui ces différents acteurs font alliance, rien ne nous empêche de penser qu'ils finiront peut-être un jour par s'opposer. Une fois que le contenu de tous les livres sera disponible sur Internet, aurons-nous encore besoin de produire des livres sur un support papier ? Car, finalement, cette double opération pour tout nouveau livre à venir a-t-elle un bon sens économique ?

Les questions que pose Google Print.

Google Print pose finalement deux séries de questions, dont les réponses sont loin d'être évidentes aujourd'hui. La première série de question concerne une éventuelle crise que vivrait l'industrie du livre. Cette crise serait loin d'être anodine puisque l'organisation du savoir dans les sociétés modernes s'est en grande partie fondée sur le livre. On entrevoit alors la deuxième série de questions qui concernent la façon de façon de faire de la recherche, c'est-à-dire la façon de consulter l'information scientifique, mais aussi de la produire.

Au sujet d'une éventuelle crise de l'industrie du livre, les questions qui se posent sont les suivantes :

- A-t-on encore besoin des bibliothèques ? Ou, plus exactement, a-t-on encore besoin de bibliothèques non virtuelles, celles où sont stockés des millions d'objets faits de papier et de couvertures cartonnées ? Car si l'information est numérisée et disponibles en n'importe quel endroit du globe, les bibliothèques non virtuelles, comme la Bnf par exemple, ne deviennent-elles pas de simples conservatoires d'objets et non plus de savoirs ?

- A-t-on encore besoin des maisons d'éditions ? Bien que les innovations en matière de fabrication de papier (à partir de bois et non plus de tissus) et d'imprimerie aient considérablement réduit le coût de fabrication des livres, la matérialité de cet objet impose des contraintes importantes (coût et capacité technique) qui limitent l'ampleur des informations qui peuvent être publiées. Il faut ainsi faire de choix. Sélectionner ce qui sera publié, et ce qui ne le sera pas. C'est la raison d'être des maisons d'éditions — et des comités de rédaction dans le cas des revues — qui vont sélectionner les informations en fonction de critères qui peuvent beaucoup varier, avec tous les problèmes de légitimer que cela pose. Si tout peut se retrouver sur Internet (ce qui a été publié et ce qui ne l'a pas été), on peut légitimement se demander si ces interfaces que sont les maisons d'éditions sont

toujours utiles ? Après tout, n'est-ce pas idéalement au chercheur de décider quelle information est pertinente pour lui ? Le moteur de recherche devenant une nouvelle interface de recherche plus objective ?

- Finalement on voit poindre la question essentielle : a-t-on encore besoin du livre ? Le livre, tout comme les bibliothèques ou les maisons d'éditions sont des solutions que les sociétés ont imaginé pour répondre à un problème : l'organisation et la diffusion de l'information au plus grand nombre et sur le long terme. Si ce problème est supprimé, pourquoi les solutions resteraient-elles en place ? Il est possible que cette remarque ne s'applique qu'à l'information scientifique et non pas au roman. Cependant, l'enjeu n'en est pas moins important.

Au sujet des pratiques de recherche, la principale question serait enfin de savoir si, avec la disparition du livre comme moyen d'organiser l'information scientifique, le chercheur sera encore amené à produire des textes linéaires avec un début, un milieu et une fin, et un principe de progression logique entre la problématique, les hypothèses, le développement et les théories. Si les pratiques de consultation de l'information changent, puisqu'elles consistent non plus à tourner des pages mais à effectuer des requêtes sur un moteur de recherche, la façon de produire l'information scientifique et de l'organiser n'est-elle pas également amenée à changer ?

Toutes ces questions sont autant d'occasions de débats qu'en ce sens le projet Google Print contribue à ouvrir pleinement. .

Résumé

Google Print, un des multiples projets lancés par l'entreprise Google, consiste à scanner et à rendre accessible sur Internet l'ensemble de l'information contenue actuellement dans les livres. Cette initiative, qui suscite de nombreuses oppositions, mais aussi des projets concurrents, est susceptible d'avoir un impact important sur les pratiques de recherche et sur l'industrie du livre. ...

Notes

1 After study and scientific research in the Us (Harvard, Ph.D. in physical chemistry), France (Cnrs, Bordeaux) and Switzerland (Eth and Epfl), he entered publishing at Elsevier Science, working for three years in the chemistry program. In 2000, he founded FontisMedia s.a., a Swiss company that specializes in scientific and technical communication. Although still active at FontisMedia, in Spring 2004 he joined the Ppur (Presses Universitaires Universitaires Romandes) based at the Epfl, as publisher of the Epfl Press (Ppur's English language imprint).

Bibliographie

Le site du projet Google Print.

L'explication du fonctionnement de « Search inside » sur le site d'Amazon.

Le site de l'Open Content Alliance.

L'annonce des poursuites judiciaires engagées par certains auteurs contre Google Print sur le site de la Guilde des Auteurs.

L'annonce des poursuites judiciaires engagées par certains éditeurs sur le site de l'Association des Editeurs Américains.

La présentation du projet Volltextsuche Online sur le site de la Börsenverein des Deutschen Buchhandels.

Jean-Noël Jeanneney, *Quand Google défie l'Europe*, Paris, Mille et une nuits, 2005.