

Sagesse en réseaux : la passion d'évaluer

Gloria ORIGGI

*Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in
information?*

*[Où est passée la sagesse que nous avons perdue
avec la connaissance ? Où est la connaissance
que vous avons perdue avec l'information ?]*

T.S. Eliot

Comment les êtres humains et les ordinateurs peuvent-ils agir collectivement de manière plus intelligente que n'importe quels individus, groupes ou ordinateurs ? Selon Gloria Origgi, le Web représente un réseau gigantesque de systèmes de hiérarchisation et d'évaluation des informations, où la *réputation* joue un rôle fondamental.

Avec la capacité remarquable du Web 2.0 de rassembler l'information répartie socialement afin d'obtenir des résultats intelligents, l'idée même d'*intelligence collective* est entrée dans une ère nouvelle. Ce qui n'était guère qu'une curiosité dans l'histoire de la pensée, à savoir que le rassemblement de nombreux jugements simples a un résultat plus complexe et plus précis que celui que donnerait un seul esprit expert, est de nos jours un *fait* majeur qui révolutionne notre manière

de penser quotidienne et de prendre des décisions, un fait qu'il faut prendre au sérieux. Les recherches de Google, les entrées de Wikipedia, les transactions de eBay sont toutes fondées sur la simple corrélation entre les données filtrées par le grand nombre et la pertinence pour nous de l'information que nous obtenons. Cependant, un souci majeur avec ces systèmes d'intelligence collective est que notre contrôle sur la manière dont l'information est traitée est quelquefois faible, et que les capacités individuelles ou institutionnelles d'intervention sur le dessein du processus de collecte d'information sont souvent très limitées. Je pense que nous devrions prendre la question du *design* très au sérieux afin d'être conscient des biais¹ de ces systèmes et des abus possibles qui peuvent en être faits. Et c'est là une question générale concernant le dessein des institutions. Peu importe le nombre de gens qui sont impliqués dans la production d'un résultat collectif – une décision, une action, une réalisation cognitive, etc. – : la manière dont leurs interactions s'effectuent, ce qu'ils peuvent savoir et ne pas savoir les uns des autres, la manière dont ils ont accès à la procédure collective, le chemin que suivent leurs actions et la manière dont il se fond dans l'action des autres, affectent le contenu du résultat. Bien sûr, cela est bien connu des décideurs politiques, des auteurs de constitutions et de tous ceux qui participent au *design* d'un système démocratique, ou de n'importe quel autre système de règles qui doit prendre en compte le point de vue du grand nombre. Mais cette prétention peut sembler moins évidente – ou du moins elle peut avoir besoin d'une justification plus explicite – lorsqu'elle concerne le *design* de la connaissance et des pratiques épistémiques sur le Web. Cela vient du fait que le Web a été principalement regardé comme une technologie révolutionnaire dont l'effet immédiat était de faire éclater toutes les procédures légitimes existantes d'accès à la connaissance en donnant ainsi à ses usagers une nouvelle liberté intellectuelle, la liberté de produire, d'avoir accès et de diffuser des contenus d'une manière totalement libre.

¹ Le terme « biais » signifie, en sciences sociales et en psychologie, qu'un système (cognitif, social, ou technologique) a des « tendances », des inclinaisons, qui ne découlent pas d'une logique purement rationnelle.

Pourtant, les méthodes pour saisir la sagesse des foules sur le Web sont nombreuses et bien plus clairement multiples qu'on ne le reconnaît ordinairement. Dans son livre sur *The Wisdom of Crowds*², James Surowiecki écrit sur les différentes manières de capter la sagesse collective. Il dit : « Finalement, il n'y a rien dans le marché des futures (*futures market*)³ qui le rende en lui-même plus brillant que, disons Google. Toutes ces tentatives reviennent à s'appuyer sur la sagesse de la foule, et c'est bien pour ça qu'elles marchent. » Cependant, le diable est parfois dans les détails et la manière dont la sagesse des foules est captée introduit une différence considérable dans le résultat et dans l'impact qu'elle a sur notre vie cognitive. La *question du design* qui est ainsi centrale lorsque l'on traite de ces systèmes est la suivante : *Comment les êtres humains et les ordinateurs peuvent-ils être reliés de telle sorte que — collectivement — ils agissent de manière plus intelligente que n'importe quels individus, groupes ou ordinateurs ?* Je vais essayer d'examiner certains des *systèmes de sagesse collective* qui sont aujourd'hui en usage sur le Web. Je vais donner une brève description « technique » de la structure informationnelle qui sous-tend chacun d'eux. Ensuite, j'avancerai l'hypothèse que ces systèmes marchent en vertu de leur manière très spéciale d'articuler d'un côté les choix individuels et les préférences filtrées collectivement, et les actions humaines et les processus informatiques de l'autre. Je conclurai ensuite par quelques remarques épistémologiques sur le rôle de la hiérarchisation dans nos pratiques épistémiques. Mon idée, c'est que le succès du Web en tant que pratique épistémique vient de sa capacité à fournir non pas tant un système potentiellement infini de stockage de l'information, qu'un réseau gigantesque de systèmes de hiérarchisation et d'évaluation dans lesquels l'information prend de la valeur pour autant qu'elle a déjà été filtrée par d'autres êtres humains. Ma modeste prévision épistémologique est que l'*Âge de l'information* est en train d'être remplacé par un *Âge de la réputation* dans lequel la réputation de quelque chose – c'est-à-dire la manière dont les autres

² Anchor, 2005. Traduction française : *La Sagesse des foules*, Jean-Claude Lattès, 2008.

³ Le marché des futures (*Futures market*) est une technique de prédiction qui consiste à faire « parier » sur le Web plusieurs acteurs sur la réalisation d'un événement futur et utiliser le résultat du pari comme prédiction.

l'évaluent et la classent — est la seule manière dont nous pouvons tirer une information à son sujet. Cette *passion de hiérarchiser* est un trait central de la sagesse collective.

James Surowiecki détermine une liste très éclairante de conditions pour la caractérisation d'une *foule sage*. Toutes les foules ne sont pas sages. Pour éviter les phénomènes bien connus tels que l'opposition de groupe, les cascades d'information, le conformisme, il faut qu'un groupe manifeste certains traits qui en font une entité potentiellement intelligente. Surowiecki propose quatre caractéristiques principales : 1/ *la diversité d'opinion* (chaque personne doit avoir une formation privée) ; 2/ *l'indépendance* (les opinions ne sont pas déterminées par les autres) ; 3/ *la décentralisation* (les êtres humains sont susceptibles de s'appuyer sur une connaissance locale) ; 4/ *l'agrégation* (la présence de mécanismes qui transforment les jugements individuels en décisions collectives). Je suis tentée d'ajouter une cinquième caractéristique qui est cruciale pour moi spécialement pour « accélérer » le filtrage collectif de l'information : 5/ *la présence d'un moyen de classement* (chaque personne devrait être susceptible de produire une hiérarchie de classement, de s'appuyer sur les systèmes de hiérarchisation passés et de faire en sorte – au moins dans certaines circonstances – que son classement soit accessible à d'autres). Je pense que cette dernière condition est particulièrement utile pour comprendre les processus d'intelligence collective que le Web 2.0. a rendus possibles, bien qu'elle ne se limite pas à cela. Bien sûr, cela ouvre la question épistémologique de la *valeur épistémique* de ces hiérarchisations, c'est-à-dire de la mesure dans laquelle leur production et leur usage dans un groupe change la proportion entre les vérités et les faussetés produites par ce groupe, et, individuellement, de la mesure dans laquelle une conscience des hiérarchisations doit affecter les croyances d'une personne. Après tout, les hiérarchisations introduisent un biais dans le jugement et la supériorité épistémique d'un jugement biaisé a besoin de justification. En outre, ces hiérarchisations sont le produit d'activités collectives humaines enregistrées par des moyens artificiels. Le contrôle de l'heuristique et des techniques qui sous-tendent

cette dynamique d'information peut ne pas être visible, voire être incompréhensible aux usagers qui se trouvent dans la position très vulnérable de s'appuyer sur des sources d'information extérieures par l'intermédiaire d'un canal dynamique de communication, utilisant des machines dont l'heuristique et les biais ne sont pas sous leur contrôle. Par exemple, le fait que des sociétés commerciales doivent ordinairement payer pour apparaître dans les moteurs de recherche ou pour être placées de manière à avoir plus de visibilité, était inconnu de 60% des usagers⁴ jusqu'à ce que l'*American Federal Trade Corporation* écrive en 2002 une recommandation publique demandant aux compagnies de moteurs de recherche de mettre en évidence leurs politiques de liens et de signaler clairement les publicités pour éviter la confusion des usagers.

Le statut épistémique de ces hiérarchisations produites collectivement ouvre ainsi une série de questions épistémologiques :

1/ Pourquoi fait-on confiance à ces hiérarchisations, et doit-on leur faire confiance ?

2/ Pourquoi devrions-nous supposer que le filtrage collectif des préférences produise des résultats plus sages sur le Web ?

3/ Quels sont l'heuristique et les biais des systèmes de collecte d'information sur le Web dont les gens devraient être conscients ?

Ces questions impliquent une perspective descriptive ainsi qu'une perspective normative sur l'épistémologie sociale des *systèmes de sagesse collective*. Une approche socio-épistémologique de ces questions – comme celle que j'adopte – devrait essayer d'élucider ces deux perspectives. Bien que je me dispose ici à explorer davantage le côté descriptif de la question, en illustrant le format des systèmes de sagesse collective avec leurs biais respectifs, j'aimerais d'abord

⁴ Princeton Survey Research Associates, « A Matter of Trust : What Users Want from Websites », Princeton, Janvier 2002 ; at <http://www.consumerWebwatch.com/news/report1.pdf>. R. Rogers évoque ce point dans *Information Politics on the Web*, MIT Press (2004).

présenter quelques réflexions épistémologiques générales qui suggèrent également une ligne possible de réponse aux questions normatives. Selon moi, dans un environnement à forte *densité informationnelle*, où les sources sont en concurrence permanente pour gagner l'attention des usagers, et où la vérification directe de l'information n'est tout simplement pas possible à des coûts raisonnables, l'évaluation et les classements sont des outils épistémiques et des pratiques cognitives qui introduisent inévitablement un raccourci dans l'information. Cela est particulièrement frappant dans les sociétés contemporaines saturées d'informations, mais je pense que c'est là un trait permanent de toute information tirée d'un corpus de connaissance. Il n'y a pas de connaissance idéale sur laquelle nous puissions nous prononcer sans avoir accès à des évaluations antérieures et faites par d'autres. Et ma modeste prévision épistémologique est que plus le contenu de l'information est incertain, plus le poids des opinions des autres pour établir la qualité de ce contenu est important. Cela ne nous rend pas plus crédules. Notre responsabilité épistémique à traiter de ces systèmes réputationnels consiste à être conscients des biais que le format de chacun ces systèmes implique, soit pour des raisons techniques, soit pour des raisons sociologiques ou institutionnelles. Une présentation détaillée des modes d'agrégation de choix individuels que l'Internet rend disponible devrait ainsi s'accompagner d'une analyse des biais éventuels que chacun de ces systèmes contient dans son *design*.

1/ Intelligence collective à partir de choix individuels

Certains agents intelligents pensent souvent mieux en groupe et d'une façon qui aurait été purement et simplement impossible à des individus isolés. L'Internet en est certainement une illustration. C'est la raison pour laquelle le développement de l'Internet a engendré dès le début des espérances considérables sur la possibilité de « surmonter » les processus de pensée au niveau individuel, vers l'émergence d'une forme nouvelle – plus puissante – d'intelligence opérant par des moyens techniques. Une pléthore d'images et de métaphores de l'Internet en tant qu'agent disposant d'une intelligence supérieure a ainsi envahi la littérature des études des nouveaux

média – telles que l’Internet en tant qu’esprit étendu, conscience digitale répartie, être intelligent d’un ordre supérieur, etc.

Cependant, les processus collectifs qui font de l’Internet un tel média cognitif puissant sont précisément un exemple d’« intelligence collective », c’est-à-dire un moyen de collecte des choix et des préférences individuels. Néanmoins, ce que l’Internet a rendu possible – cela a été effectivement spectaculaire – a été une toute nouvelle forme de collecte qui n’existait tout simplement pas avant son invention et sa diffusion mondiale. En ce sens, il a donné un nouvel outil pour collecter des comportements individuels qui sont susceptibles de servir de base pour repenser d’autres formes d’institutions dont la survie est liée à la combinaison appropriée des opinions du grand nombre.

1.1 L’Internet et le Web

L’aspect frappant de cette nouvelle forme d’agrégation est une manière spéciale d’articuler les choix individuels et les préférences filtrées de manière collective par l’intermédiaire de la technologie de l’Internet et, spécialement, du *World Wide Web*. En ce sens, il est utile de faire une distinction claire entre l’Internet en tant que phénomène de réseau et le Web en tant que technologie précise rendue possible par l’existence de ce nouveau réseau. L’Internet est un réseau dont les débuts remontent aux années soixante, lorsque les savants américains de AT & T, Rand et MIT et la *Defense Communication Agency* ont commencé à penser un autre modèle de transmission de l’information par un réseau. Dans le système téléphonique classique, lorsque vous appelez New York depuis votre appartement parisien, un circuit s’établit entre vous et votre destination à New York – en gros un fil de cuivre qui relie physiquement les deux destinations. L’idée fut ainsi de développer une autre technologie – une technologie de « commutation de paquets d’information », en digitalisant les conversations – c’est-à-dire en transformant les ondes en bits, puis en découpant le résultat en paquets qui pourraient être emportés dans un flux

indépendant par l'intermédiaire d'un réseau tout en donnant l'impression d'une connexion en temps réel de l'autre côté. Au début des années soixante-dix, le premier réseau décentralisé, Arpanet, fut créé, qui permettait de transmettre un message en diffusant ses plus grandes parties par le réseau et ensuite en le reconstruisant à l'autre bout. Au milieu des années soixante-dix, la première application importante au réseau fut créée, le mail. Ce qui a fait de ce réseau un outil aussi puissant était sa forme de croissance décentralisée : Internet est un réseau de réseaux, qui utilise les connexions préexistantes (comme les réseaux téléphoniques) pour faire communiquer des ordinateurs suivant un certain nombre de protocoles (des choses comme IP/TCP) qui ne sont la propriété de personne : chaque nouvel usager peut se connecter au réseau en utilisant ces protocoles. Chaque invention d'une application, d'un système de courrier, d'un système de transfert de vidéo, d'un système téléphonique digital, peut utiliser les mêmes protocoles. Les protocoles de l'Internet sont « communs »⁵ et ce fut un stimulant à la croissance du réseau et à la créativité des applications qui l'utilisent. C'est là un point crucial pour la sagesse du net. Sans le choix politique du maintien du libre accès à ces protocoles, le net n'aurait pas crû d'une manière décentralisée et les pratiques de connaissance en collaboration qu'il a permis de réaliser n'auraient pas été possibles.

Le *World Wide Web*, qui est une invention bien plus récente, a conservé la même philosophie de protocoles ouverts compatibles avec l'Internet (comme *http – hypertext transfer protocol* – ou *HTML – hypertext markup language*). Le Web est un service qui opère au moyen de l'Internet, un ensemble de protocoles et de conventions qui permet à des « pages » (c'est-à-dire un format particulier d'information qui facilite l'écriture et la lecture de contenus) d'être facilement reliées les unes aux autres, par la technique de l'hypertexte. C'est un protocole de visualisation qui rend très simple l'affichage de l'information. La croissance du Web n'est pas la même chose que la croissance de l'Internet. Ce qui a fait croître si vite le

⁵ Cf. sur ce point L. Lessig (2001), *The Future of Ideas*, Vintage, New York.

Web est le fait que la création d'un lien d'une page à l'autre n'exige aucune compétence technique. Le Web est une illustration de la manière dont une application d'Internet peut fleurir grâce à la libre disponibilité des protocoles. Et il est important de souligner que l'impact des TIC (Technologie de l'information et de la communication) sur l'intelligence collective vient surtout du Web.

1.2 Le Web, la mémoire collective et la méta-mémoire

Qu'est-ce qui rend la collecte des préférences individuelles par le Web tellement spéciale ? Pour l'histoire de la culture, le Web est une révolution majeure dans le stockage, la diffusion et la conservation de l'information. Les plus grandes révolutions culturelles dans l'histoire de la culture ont eu un effet sur la distribution de la mémoire. Le Web est une révolution de ce genre. Cette révolution a souvent été comparée à l'invention de l'écriture ou de l'imprimerie. Les deux comparaisons sont valables.

L'écriture, apparue à la fin du quatrième millénaire avant Jésus-Christ en Mésopotamie, est un support de mémoire externe qui rend possible une réorganisation de la vie intellectuelle et une structuration des pensées, non-accessibles dans les cultures orales. Avec l'apparition de l'écriture, une partie de notre cognition « quitte » le cerveau pour être déposée dans des supports extérieurs. La représentation visuelle de la connaissance d'une société rend possible à la fois de réorganiser la connaissance d'une manière plus utile, plus « logique », en utilisant, par exemple, des listes, des tables ou des arbres généalogiques, et de la fixer d'une génération à l'autre. De surcroît, la naissance de castes « managériales » qui supervisent la mémoire culturelle, comme les scribes, les astrologues et les bibliothécaires, rend possible l'organisation d'une méta-mémoire, c'est-à-dire l'ensemble des processus permettant d'accéder et de retrouver la mémoire culturelle.

L'imprimerie, apparue dans notre civilisation à la fin du XV^e siècle, redistribue la mémoire culturelle, en changeant la configuration de la « pyramide

informationnelle » dans la diffusion de la connaissance. En quel sens la révolution du Web est-elle comparable à l'invention de l'écriture et de l'imprimerie ? Dans la lignée de ces deux révolutions précédentes, le Web augmente l'efficacité dans l'enregistrement, la sauvegarde, la reproduction et la répartition de la mémoire culturelle. Comme l'écriture, le Web est un support de mémoire extérieur, mais elle s'en distingue en ce qu'elle est « active » par opposition à la nature passive de l'écriture. Comme l'imprimerie, le Web est un moyen de distribuer la mémoire culturelle dans une population, mais il en diffère de manière importante dans la mesure où il modifie de manière cruciale les coûts et le temps de la distribution. Mais à la différence de l'écriture et de l'imprimerie, le Web rend possible un changement radical dans les conditions d'accès et de sauvegarde de la mémoire culturelle avec l'introduction de nouveaux moyens de gérer la méta-mémoire, c'est-à-dire les *processus par lesquels on accède à la mémoire*.

La civilisation, dans une large mesure, consiste en la conception, l'organisation et l'institutionnalisation d'une méta-mémoire efficace, c'est-à-dire d'un système de règles, de pratiques et de représentations qui nous permettent de nous orienter utilement dans la mémoire collective. Une bonne partie de notre éducation scolaire consiste à assimiler des systèmes de méta-mémoire, des classifications de style, des hiérarchisations, etc., choisis par notre culture particulière. Par exemple, il est important de connaître les bases de la rhétorique afin de « classer » rapidement un vers comme appartenant à un certain style, et par suite à une certaine période, afin d'être en mesure de le situer efficacement dans le corpus de la littérature italienne. Ainsi, la méta-mémoire n'a pas seulement une fonction cognitive – retrouver une information dans un corpus –, mais aussi la fonction sociale et épistémique de fournir une organisation à cette information en termes de systèmes différents de classifications qui incorporent la valeur de la « tradition culturelle » de ce corpus. La manière dont nous récupérons de l'information est une activité épistémique qui nous permet d'avoir accès, au moyen des filtres de triage, à la manière dont les autorités culturelles ont classé et rangé une information à

l'intérieur de ce corpus. Avec l'apparition de technologies qui automatisent les fonctions d'accès à la mémoire, comme les moteurs de recherche et les systèmes de traitement de la connaissance, la méta-mémoire devient également une partie de la mémoire extérieure : une fonction cognitive, centrale à l'organisation culturelle des sociétés humaines, est devenue automatisée — un autre « morceau » de la cognition quitte ainsi notre cerveau pour se matérialiser en des supports extérieurs. Pour revenir à l'exemple évoqué plus haut, si j'ai à l'esprit un vers de poésie, disons « *Guido, i'vorrei...* », mais que je ne puis me rappeler ni de l'auteur ni de la période, et que je ne suis pas capable d'identifier le style, je peux aujourd'hui tout simplement écrire ce vers dans la fenêtre d'un moteur de recherche et regarder les résultats. La combinaison hautement improbable de mots dans un vers rend possible un choix suffisamment pertinent d'informations qui donne parmi les premiers résultats le poème dont le vers est tiré (ma recherche sur ce vers a donné avec Google 654 réponses, dont les dix premières contenaient le texte complet du poème de Dante).

Comment la technologie du Web conçoit-elle cette méta-mémoire ? Ce qui est exceptionnel en ce qui concerne le Web, c'est que les actions des usagers laissent une trace sur le système, qu'il peut immédiatement réutiliser, comme les traces que laissent les escargots sur le sol, qui manifestent aux autres escargots le chemin qu'ils suivent. La combinaison des traces des différents types d'usage peut être aisément disposée en un classement qui informera et influencera les préférences et les actions futures des usagers. Le corpus de connaissance accessible sur le Web – construit et conservé par les comportements individuels des usagers – est automatiquement filtré par des systèmes qui collectent ces comportements dans un classement qui les rend accessibles en tant qu'information filtrée à de nouveaux usagers individuels. J'analyserai deux différentes classes de systèmes méta-mémoriels. Ces systèmes, bien qu'ils fournissent tous les deux une sélection d'information qui à la fois informe et influence le comportement des usagers, sont conçus d'une manière différente, une différence qu'il ne faut pas négliger.

2. Le filtrage collaboratif : la sagesse à partir d'algorithmes

2.1. Les systèmes de traitement de la connaissance

Le filtrage collaboratif est une manière de faire des prévisions sur les préférences d'un usager fondée sur le type de comportement de beaucoup d'autres usagers. Il est principalement utilisé à des fins commerciales dans les applications du Web pour le e-business, bien qu'il ait été étendu à d'autres domaines. Un exemple bien connu de système de filtrage collaboratif dont nous sommes tous familiers, est le site Amazon.com : Amazon.com est une application du Web, un système de traitement de la connaissance qui conserve la trace des interactions des usagers avec le système et qui est destiné à manifester les corrélations entre les types d'activités d'une manière qui informe les usagers des préférences des autres usagers. Le trait le mieux connu de ce système est celui qui relie différentes marchandises à acheter : « les clients qui ont acheté X achètent aussi Y ». L'originalité de ces systèmes est que le couplage de X et de Y s'effectue en un sens *de la base au sommet* (bien que la valeur des seuils d'activités au-dessus desquelles cette corrélation apparaît soit fixée par l'architecture de l'information du système). La corrélation entre le livre de James Surowiecki et le livre de Ian Ayer *Super Crunchers*, que vous pouvez trouver sur la page d'Amazon pour le livre *The Wisdom of Crowds*, a été produite automatiquement par un algorithme qui rapproche les préférences des usagers et fait apparaître la corrélation. C'est là un trait caractéristique de ces systèmes interactifs, dans lesquels de nouvelles catégories sont créées en transformant automatiquement des actions humaines en classements visibles. La sagesse collective du système vient d'une division du travail cognitif entre les algorithmes qui composent et visualisent l'information, et les usagers qui interagissent avec le système. Les classifications et les hiérarchisations qui sont ainsi créées ne sont pas fondées sur une connaissance culturelle antérieure des us et coutumes des usagers, mais sur l'apparition d'agencements significatifs de préférences collectées à travers les interactions individuelles avec le système. Bien sûr, les biais sont possibles dans le système : les poids accordés à chaque article qui apparaît dans une corrélation sont fixés de telle

sorte que certains articles ont plus de chances d'être recommandés que d'autres. Mais étant donné que le système est alimenté par les actions répétées des usagers, une recommandation trop biaisée qui rapprocherait des articles que les usagers n'achèteront pas ensemble ne se reproduira pas assez de fois pour se stabiliser à l'intérieur du système.

2.2. Le *PageRank*^a

Une autre classe de systèmes qui réalisent des fonctions de méta-mémoire par des moyens artificiels rassemble les moteurs de recherche. Comme nous le savons tous par expérience, les moteurs de recherche ont constitué une transformation majeure de nos pratiques épistémiques et une révolution cognitive profonde. L'innovation la plus remarquable de ces outils vient de la découverte de la structure du Web au commencement de ce siècle⁶. La structure du Web est celle d'un réseau social, et elle contient beaucoup d'information sur les préférences et les habitudes de ses usagers. Les moteurs de recherche de deuxième génération, comme Google, sont capables d'exploiter cette structure pour obtenir de l'information sur la manière dont la connaissance est répartie dans le monde entier. Fondamentalement, l'algorithme de *PageRank* interprète un lien d'une page A à la page B comme un vote que la page A exprime au sujet de la page B. Mais sur le Web nous ne sommes pas en démocratie et les votes n'ont pas tous le même poids. Les votes qui viennent de certains sites – appelés « *hubs* » – ont bien plus de poids que d'autres, et ils reflètent en un sens des hiérarchies de réputation qui existent en dehors du Web. En gros, un lien depuis la page d'accueil de mon site à la page de l'Université de Harvard pèse bien moins lourd qu'un lien de la page de l'université de Harvard à ma propre page d'accueil. Le Web est un réseau « aristocratique » – une expression utilisée par les théoriciens des réseaux sociaux – c'est-à-dire un réseau dans lequel « les riches deviennent plus riches » et dans lequel plus vous recevez de liens, plus haute est la probabilité que

^a Le *PageRank* ou *PR* est le système de classement des pages Web utilisé par le moteur de recherche Google pour attribuer l'ordre des liens dans les résultats de recherche.

⁶ Kleinberg, J. (2001) « The Structure of the Web », *Science*.

vous en recevrez davantage encore. Cette disparité crée un « paysage réputationnel » qui influence le résultat d'une recherche. L'algorithme de *PageRank* est alimenté par la connaissance et les préférences locales de chaque usager individuel et il les influence en affichant une hiérarchie de résultats qui sont interprétés comme une hiérarchie dans la pertinence. Remarquez que ce système *n'est pas* un système de traitement de la connaissance : l'algorithme de *PageRank* ne connaît rien des activités particulières de chaque individu : il ne sait pas combien de fois vous et moi allons sur le site JSTOR et il ne fait pas de rapprochements entre nos navigations. Un « click » d'une page vers une autre est une information obscure pour *PageRank*, tandis qu'un lien entre deux pages contient beaucoup d'information sur la connaissance des usagers que le système est en mesure d'extraire. Pourtant, les deux systèmes sont comparables du point de vue de la forme de l'intelligence collective : ni l'un ni l'autre n'exigent aucune coopération entre des agents pour créer un système de partage des choix. L'aspect « de collaboration » du filtrage collectif est davantage aux mains de machines que d'agents humains⁷. Le système exploite l'information que des agents humains, soit laissent à leur insu sur le site web en interagissant avec lui (systèmes de traitement de la connaissance), soit produisent de manière active en faisant un lien d'une page à une autre (les moteurs de recherche) : le résultat en est collectif, mais la motivation en est individuelle.

Les biais des moteurs de recherche ont fait l'objet ces dernières années de bien des discussions, des controverses et des craintes collectives. Comme je l'ai mentionné, l'amélioration des moteurs de recherche de deuxième génération comme Google a permis au moins de marquer explicitement les inclusions et les placements privilégiés, mais cela a nécessité une intervention politique. De même, l'« effet Matthew » des réseaux aristocratiques n'est que trop connu, et le risque de ces outils est de donner la prééminence à des sites déjà puissants aux dépens des autres. La

⁷ Les systèmes de traitement de la connaissance comme Amazon.com ont certains traits de filtrage en collaboration qui ont besoin de coopération, comme le fait d'écrire le compte rendu d'un livre avec le système de classement à cinq étoiles, mais ces traits ne sont pas essentiels au fonctionnement du processus de filtrage en collaboration.

conscience de ces biais devrait entraîner également une amélioration des pratiques de recherche : par exemple : le plus improbable est l'enchaînement des mots-clés, le plus correcte est le résultat filtré. On devrait donner aux novices et aux apprentis des principes simples semblables qui les rendront moins vulnérables à ces biais.

3. Les systèmes de réputation : la sagesse à partir de l'angoisse du statut

Le filtrage collaboratif de l'information peut exiger quelquefois une participation plus active à une communauté que cela n'est nécessaire dans les exemples évoqués plus haut. Dans son ouvrage sur les politiques informationnelles du Web (*Information Politics on the Web*), le sociologue Richard Rogers classe la dynamique du Web comme « volontaire » ou « non-volontaire » selon le rôle respectif que jouent les êtres humains et les machines dans le *feedback* d'information qu'ils donnent aux usagers. Les *systèmes de réputation* sont un exemple d'une application plus « volontaire » sur le Web que les applications évoquées plus haut. Un système de réputation est un genre spécial d'algorithme de filtrage en collaboration qui détermine les classements pour un ensemble d'agents fondés sur les opinions que ces agents ont les uns des autres. Un système de réputation collecte, répartit et rassemble les réponses [*feedback*] concernant le comportement passé des participants.

Le système de réputation le mieux connu et probablement le plus simple qui ait une forte influence sur le Web est le système de vente aux enchères sur le site www.eBay.com. EBay permet des interactions commerciales entre plus de 125 millions de personnes dans le monde. On y est acheteur ou vendeur. Les acheteurs font une offre pour un article. Si leur offre est acceptée, ils font l'opération commerciale, puis les uns et les autres (acheteurs et vendeurs) font une évaluation de la qualité de cette transaction. Les différentes évaluations sont ensuite rassemblées par le système dans un profil très simple d'évaluation, où les évaluations positives et les évaluations négatives avec quelques mots de commentaire sont affichées sur l'écran pour les usagers. La réputation d'un agent est ainsi une information utile pour

se décider à poursuivre une transaction. Dans ce cas, la réputation a une valeur réelle, mesurable, commerciale : sur un marché d'une offre limitée avec très peu d'information disponible pour chaque offre, la réputation devient une information cruciale pour faire confiance au vendeur. Les vendeurs sur eBay savent très bien la valeur de leur réputation dans un tel contexte commercial (pas de rencontres physiques, aucune possibilité de voir ni de toucher l'article, flou sur le cadre normatif de la transaction — si par exemple elle a lieu entre deux pays différents, etc.), aussi y a-t-il un grand nombre de transaction à prix très bas dont l'objectif est précisément d'acquérir une évaluation plus positive. Le système crée un résultat collectif forçant à la coopération, c'est-à-dire en demandant aux usagers de laisser une évaluation à la fin de la transaction et en les sanctionnant s'ils ne s'y conforment pas. Sans cette participation active des usagers, le système serait inutile. Pourtant, c'est là une forme spéciale de comportement en collaboration qui n'exige aucune adhésion à la coopération en tant que valeur. Les usagers non coopératifs sont sanctionnés à différents niveaux : ils peuvent recevoir une évaluation négative non seulement si la transaction n'est pas saine, mais également s'ils ne prennent pas part au processus d'évaluation. Violer les règles sur eBay peut conduire à l'exclusion de la communauté. Cette forme de sagesse implique ainsi une participation active de la part des usagers par crainte d'être ostracisé par la communauté (ce qui serait perçu comme une perte d'occasions commerciales). Des biais sont ici aussi clairement possibles. Certains font des transactions à bas prix dans le seul but d'acquérir des points de réputation. C'est là un biais dont on doit être conscient et que l'on peut aisément vérifier : si un vendeur propose trop d'articles à bon marché, il se soucie bien trop de son image publique pour être considéré comme digne de confiance.

Certains traits réputationnels sont également utilisés par des systèmes non commerciaux comme www.flickr.com. Flickr est une plate-forme de collaboration de partage d'images. Pour chaque image, vous pouvez visualiser le nombre d'usagers qui l'ont placée parmi leurs images favorites et qui ils sont.

Les systèmes de réputation diffèrent d'autres systèmes de mesure de la réputation qui utilisent l'analyse des citations, comme par exemple, le *Science Citation Index*. Ces systèmes sont en un sens fondés sur la réputation, étant donné qu'ils utilisent des techniques sciento-métriques pour mesurer l'impact d'une publication en fonction du nombre de fois où elle est citée dans d'autres publications. Mais ils n'exigent aucune participation active des agents pour obtenir la mesure de la réputation.

4. Systèmes de collaboration, systèmes ouverts : la sagesse par la coopération

Le filtrage collaboratif sur le Web peut être encore plus volontaire et fondé sur l'homme que dans les exemples précédents, tout en nécessitant encore un support Web pour produire un résultat intelligent. Deux d'entre eux sont les systèmes de collaboration les plus discutés, qui doivent leur succès à une coopération humaine active dans le filtrage et la révision de l'information rendue disponible : les communautés de développement de logiciels *Open Source*, comme Linux, et les projets de contenu collaboratif ouvert comme Wikipedia. Dans ces deux cas, le processus de filtrage est entièrement fait de main d'homme : le code ou le contenu est accessible à une communauté qui peut le filtrer pour le corriger, le rédiger ou l'effacer selon des critères de qualité personnels ou partagés. Je dirais que ce sont là des communautés d'*amateurs* plutôt que d'experts, c'est-à-dire de gens qui aiment ce qu'ils font et qui choisissent de partager leur connaissance au profit de la communauté. Une sagesse collective est ainsi créée par les efforts humains individuels qui sont rassemblés dans une entreprise commune dans laquelle certaines normes de coopération sont partagées.

Je ne discuterai pas des biais de Wikipedia : c'est un sujet tellement large qu'il pourrait faire l'objet d'une autre étude. On pourrait seulement mentionner que Larry Sanger, l'un de ses fondateurs, est en train de promouvoir un autre projet, www.citizendum.org qui adopte une politique d'accréditation de ses auteurs. Bien sûr, l'auto-promotion, l'idéologie, les attaques ciblées de la réputation de quelqu'un,

peuvent fonctionner comme des biais dans le choix des entrées. Mais les faits ont montré que la crainte de Wikipedia en tant que lieu dangereux d'information tendancieuse est infondée : grâce à sa taille importante, Wikipedia connaît une diversité considérable de sujets et d'opinions, et l'on a montré que sa fiabilité n'était pas moindre que celle de l'*Encyclopedia Britannica*⁸.

5. Les systèmes de recommandation : la sagesse à partir des connaisseurs

Une autre classe de systèmes se fonde sur les recommandations de connaisseurs dans un domaine particulier. Un exemple de sagesse créée à partir de recommandations d'experts est le *Music Genoma Project* [Projet de Génôme Musical] sur le site www.pandora.com, une espèce de radio basée sur le Web qui fonctionne en rassemblant des milliers de descriptions et de classifications de morceaux de musique écrites par des connaisseurs et qui associe ces descriptions aux « goûts » d'auditeurs (comme ils disent). Ensuite, elle diffuse un choix de morceaux de musique qui correspond à ce que les auditeurs aiment écouter. Un autre exemple de système de recommandation est www.CiteUlike.org, un service gratuit en ligne permettant de partager les textes que vous lisez avec d'autres membres de votre communauté scientifique, de trouver qui est en train de lire le même texte, d'organiser une bibliothèque de certains des textes que vous lisez en les étiquetant d'une certaine manière. Ici, les préférences sont volontairement partagées. Certains systèmes de recommandation collectent de l'information chez les usagers en leur demandant activement de classer un certain nombre d'articles, ou d'exprimer une préférence entre deux d'entre eux, ou de créer une liste d'articles qu'ils aiment. Le système compare ensuite les données à d'autres données semblables collectées chez d'autres usagers et affiche la recommandation. C'est fondamentalement une technique de filtrage collaboratif avec une composante plus active : on demande aux gens d'exprimer leurs préférences, au lieu de se contenter d'inférer leurs préférences à partir de leurs comportements, ce qui fait une différence considérable : on sait fort

⁸ Cf. « Internet Encyclopaedias go head to head », *Nature*, 438, 15 décembre 2005.

bien en psychologie que nous ne sommes pas très bons en introspection et que parfois nous exprimons consciemment des préférences qui ne s'accordent pas avec notre comportement : si on me le demande, je puis exprimer une préférence pour la musique classique, alors que si je note le nombre de fois par semaine où j'écoute de la musique classique par rapport à d'autres genres de musique, je me rends compte que mes préférences sont tout à fait autres.

Conclusions

Cette longue énumération d'outils sociaux du Web produisant une sagesse collective illustre à quel point le type de rassemblement de choix et de préférences individuels peut être diversifié. Les différences de format que j'ai soulignées aboutissent à de profondes différences dans le genre de communautés collectives qui sont engendrées par le ITC. Quelquefois, la communauté est absente, comme dans le cas des usagers de Google, qui ne peuvent être définis comme une « communauté » en aucun sens normatif intéressant ; quelquefois, la communauté est exigeante en matière de normes comme dans le cas de eBay, où la participation au processus de filtrage est nécessaire à la survie de la communauté. Si la nouvelle production collective de connaissance que le Web – et en particulier le Web 2.0. et 3.0 – rend possible devait servir de laboratoire pour inventer de « meilleures » procédures collectives de production de connaissance ou de décisions sages, ces différences devraient être prises en compte.

Mais permettez-moi de revenir pour conclure sur une exigence plus épistémologique concernant le genre de connaissance produit par ces nouveaux outils. Comme je l'ai dit au début, ces outils marchent tant qu'ils donnent accès à des hiérarchisations d'information, à des labellisations de procédures et d'évaluations. Même Wikipedia, qui n'affiche aucun moyen explicite de classement, fonctionne sur le principe suivant : si une nouvelle entrée a survécu sur le site – c'est-à-dire, si elle n'a pas été effacée par d'autres wikipédiens – elle mérite d'être lue. Cela peut être un outil d'évaluation trop faible, et, comme je l'ai dit, la discussion se poursuit

aujourd'hui sur l'opportunité d'introduire des moyens de filtrage plus structurés dans Wikipedia⁹, mais mon opinion est que la survie de projets égalitaristes comme Wikipedia dépend de leur capacité à intégrer un classement : le label Wikipedia en lui-même fonctionne déjà comme un signal réputationnel qui oriente le choix des usagers. Sans la réputation du label, le succès du projet serait bien plus limité.

Le Web n'est pas seulement un puissant réservoir de toutes sortes d'information labellisée ou non labellisée, mais elle est également un puissant outil *réputationnel* qui hiérarchise, introduit des systèmes de classement, de poids et de biais dans le paysage de la connaissance. Même dans ce monde dense en information, une connaissance sans évaluation serait un triste paysage désertique dans lequel on serait sidéré devant une masse énorme et muette d'information, comme Bouvard et Pécuchet, les deux héros du célèbre roman de Flaubert, qui choisirent de se retirer pour apprendre toutes les disciplines connues pour en fin de compte ne pouvoir rien apprendre. Un système efficace de connaissance se développera inévitablement en engendrant une diversité d'outils évaluatifs : c'est ainsi que la civilisation se développe, que des traditions se créent. Une tradition culturelle est pour commencer un système d'identification de ceux qui sont dedans et de ceux qui sont dehors, de ceux qui émergent et de ceux qui sont noyés dans le magma du passé. La bonne nouvelle est qu'à notre époque, cette évaluation inévitable s'effectue par des outils nouveaux, collectifs qui remettent en question les idées reçues et développent et améliorent une manière innovante et démocratique de filtrer la connaissance. Mais on ne saurait échapper à la création d'un corpus de connaissance « canonique » – même à chaque fois provisoire et révisable.

⁹ Voir I. Sanger, « Qui dit que nous savons. Sur la nouvelle politique de la connaissance », sur www.edge.org et ma réponse, G. Origi, « Why reputation matters ».

Bibliographie

- A. Clarck (2003) *Natural Born Cyborgs*, Oxford University Press
- L. Lessig (2001), *The Future of Ideas*, Vintage, New York
- G. Origgi (2007), « Wine epistemology: The role or reputation and rating systems in the world of wine », in B. Smith (éd.) *Questions of Taste*, Oxford University Press
- G. Origgi (2007) « Un certain regard. Pour une épistémologie de la réputation », présenté à l'atelier *La réputation*, Fondazione Olivetti, Rome, avril 2007.
- G. Origgi (2008), *Qu'est-ce que la confiance ?* Vrin, Paris.
- R. Rogers (2004) *Information Politics of the Web*, MIT Press.
- L. Sanger (2007) « Who says we know: On the new Politics of knowledge », www.edge.org
- Taraborelli, D. (2008) « How the Web is changing the way we trust », in K. Waelbers, A. Briggie, P. Brey (éds.), *Current Issues in Computing and Philosophy*, IOS Press, Amsterdam, 2008.
- P. Thagard (2001), « Internet epistemology : Contributions of new information technologies to scientific research », in K. Crowley, C.D. Schunn et T. Okada (éds.), *Designing for science: Implications from professional, instructional and everyday science*, Mahwah, NJ : Erlbaum, 465-485.

Texte paru dans laviedesidees.fr, le 30 septembre 2008

© laviedesidees.fr