
Vu du web.

L'émergence du partage d'information, dynamique sociale de l'exploration du web

*Vous avez choisi vos amis,
vous vous êtes connecté aux pages auxquelles
vous vouliez vous connecter et c'est vous qui êtes le
mieux placé pour décider des choses
qui vous intéressent.*
Greg Marra¹

Disons-le d'emblée : le problème du web, c'est qu'il y a trop de choses.

On y trouve aussi bien des photos scannées du siècle passé, les performances de la bourse en temps réel ou les prévisions météo du week-end à venir. Aussi bien des tableaux de données économiques validées par une institution, que des vidéos « lolcats » postées par un inconnu ou des cartes agrégeant des sources variées suite à l'investigation de journalistes. Aussi bien des discussions cacophoniques sur des forums, que des récits à une voix déposés anonymement sur un blog ou des tribunes militantes signées par un collectif.

Il y a « trop » de « choses » : « choses » renvoie à l'hétérogénéité des documents numériques, « trop » à ce qu'un internaute peut absorber. Explorer le web est une tâche herculéenne, qui nécessite de mettre de l'ordre dans ces « choses » sans perdre leur diversité, et d'équiper l'internaute pour qu'il se retrouve dans ce « trop ». Ce chapitre retrace l'histoire de la conception des dispositifs par lesquels un internaute navigue dans les contenus numériques, une loupe sur ces boutons et balises sur lesquels tout un chacun peut cliquer mais que chacun saisit à sa manière. Le partage d'information, issu de cette généalogie, apparaît comme une solution sociale permettant d'explorer le « trop de choses » web.

Comme le montrent les exemples précédents, les contenus numériques ont différentes dates, différents formats, et différents cadres d'interaction. La perception d'un immense

¹ Greg Marra, en charge du EdgeRank, cité par Libération
(http://ecrans.liberation.fr/ecrans/2014/12/16/facebook-l-homme-qui-fait-le-mur_1165096, consulté le 05/05/2015).

bazar en ligne (Benkler, 2009) tient d'une augmentation réelle du volume de documents¹, mais aussi de l'impossibilité d'appliquer les prises traditionnellement utilisées pour se saisir d'un contenu. La référence à l'auteur, la place sur une étagère, ou la fraîcheur d'une publication ne constituent plus des méta-informations stables pour évaluer la valeur d'un document. Le numérique incorpore une volumineuse quantité de contenus, et brouille les repères qui permettent d'appréhender ces objets. Le paysage du web est à la fois plus étendu et plus flou qu'un kiosque à journaux ou une bibliothèque.

De surcroît, le numérique individualise l'exploration du web et laisse l'internaute livré à lui-même dans ce paysage. Internet est un espace où les documents sont posés à une adresse par les producteurs où c'est aux récepteurs de les visiter. Le réseau inverse ainsi le sens du média : c'est l'internaute qui va *tirer* les contenus qu'il consulte (en mode *pull*), plutôt qu'un éditeur les lui *pousse* (en mode *push*)². Devenu autonome dans sa sélection de contenus, l'internaute est responsable de sa navigation à travers ses clics. Le web déporte la charge de programmation incarnée par la navigation sur l'usager (Patriarche, 2004). De ce fait, les pratiques informationnelles en ligne sont tout autant déterminantes qu'individuelles.

Se perdre et se retrouver dans le « trop de choses » du web est donc une activité personnelle qui s'appuie sur les méta-informations publiques des contenus et sur l'équipement cognitif des internautes. Pour s'orienter en ligne, chacun doit répondre à deux questions : (1) comment s'assurer de la qualité des contenus et déterminer la hiérarchie, ou l'ordre, des documents ? (2) Par quel chemin matériel et cognitif accéder à ces contenus et poursuivre sa navigation ? Les artefacts des dispositifs accompagnent l'usager dans sa navigation, qui matérialise la réponse aux deux questions.

Les dispositifs sont ainsi, aux différents âges du web, une part du problème et une part de la solution de la surabondance informationnelle. Les artefacts techniques affichent autour des liens, des notes, un avatar, des étoiles, un commentaire ... autant de signes qui donnent des prises à l'appropriation des contenus et deviennent les outils de l'exploration individualisée du web des documents. Or ces dispositifs sont produits dans un contexte, à un certain stade du développement du réseau, par des individus ou des entreprises qui ont une certaine conception du web. Ces artisans projettent des usages et une cible d'utilisateurs, et implémentent ces projections dans des algorithmes et des interfaces nécessairement non neutres (Winner, 1980). Les internautes se saisissent ou non de ces entrées, en en percevant ou non les présupposés. Les services du web

¹ J'appellerai « document numérique » tout objet étant identifié avec une url. Pour une discussion sur la nature de ces documents par rapport aux documents traditionnels, voir par exemple Boullier & Ghitalla (2004), ou Ertzscheid (2009), ou encore les travaux de l'A.D.B.S.

² Cette remarque théorique dans le cadre de cette approche par les dispositifs devra bien sûr être nuancée en étudiant les pratiques.

méritent donc d'être étudiés pour identifier les ressorts qu'ils mettent en œuvre, en tenant compte du contexte de production de ces artefacts.

L'évolution des services au fil du temps résulte de deux leviers : les innovations techniques et la généralisation des usages. Pour le premier levier, il s'agit par exemple d'afficher le nombre de vues d'un document, indicateur qu'il n'était pas possible de produire sans une traçabilité des usages et un calcul informatique agrégeant ces traces en temps réel. Les innovations techniques héritent des fondements de l'infrastructure réseau et de la micro-informatique, et testent des usages dans les *devices*, applications, services, etc. En même temps que les innovations techniques se concrétisent (Simondon, 1958), les pratiques des internautes se sédimentent. Ainsi, la mention « @ », qui permet d'adresser une personne en public, traverse les dispositifs puisqu'elle apparaît dans les forums dès 1993 et se retrouve dans les tweets de 2015 ; alors que d'autres pratiques sont purement et simplement oubliées. La généralisation des usages constitue le deuxième moteur de développement, en s'appuyant sur les compétences des utilisateurs. Si les pionniers du web des années 1990 revendiquaient une habileté technique, ce n'est plus le cas du grand public qui accède aujourd'hui massivement au web. La généralisation de l'accès à Internet implique donc une évolution des outils vers une simplification des usages, une automatisation de certaines fonctions.

La perspective historique adoptée dans ce chapitre permet alors de rendre compte d'une certaine progression des services de l'exploration du web, progression à la fois technique et fonctionnelle. Dans une première partie, une archéologie du web retracera le développement des deux familles d'artefact visant à outiller l'exploration numérique : la première cherche à hiérarchiser les contenus ; la seconde vise à équiper la navigation des internautes sur le web. Ces dispositifs se déploient dans le web des documents, mais force est de constater que l'ambition de permettre à des internautes toujours plus nombreux d'explorer un web toujours plus grand reste utopique. La deuxième partie envisagera donc une autre famille de solutions introduite en faisant un détour par le web des sociabilités : les outils d'expression et d'interaction en ligne produisent un flux actualisé de contenus numériques qui alimente l'internaute. Facebook sera étudié plus précisément pour montrer l'imbrication des contenus dans les interactions. La pratique du partage d'information naît ainsi d'une lignée de dispositifs propres au web des documents et d'une tradition d'usages du web des interactions : il résout l'exploration du web grâce à la médiation du réseau social.

1.1) Ordre et accès dans le web des documents

Pour explorer le « trop de choses » en ligne, l'internaute doit définir ce qu'il veut consulter, une hiérarchie des contenus accessibles ; et il doit accéder à ces contenus, par sa navigation. Pour expliciter les notions d'ordre et d'accès dans le web des documents, prenons l'exemple d'un voyage. Comment visiter une destination ? Les *tour operators* peuvent proposer des voyages organisés où le programme est fixé pour tous, comme les médias de masse proposent un flux unique de contenus. Mais la navigation en ligne est un voyage où l'internaute doit lui-même construire son programme, aller au Mont Saint-Michel ou suivre les traces de Flaubert ; et aller au Mont-Saint-Michel avec sa voiture

n'est pas la même expérience qu'y aller en traversant la baie à pied avec un groupe s'intéressant à la sauvegarde de la flore. Il faut donc à la fois choisir les lieux où se rendre en les hiérarchisant, et choisir les chemins qui y mènent. La hiérarchisation des sites et l'itinéraire sont indissociables pour produire l'expérience du voyage.

Choisir les lieux à voir correspond à créer un ordre de ces lieux, en agrégeant différents critères d'évaluation. Ces critères sont fournis par les conseils de l'office du tourisme, les étoiles du guide Vert, ou une émission vue à la télévision. Parmi les outils destinés à aider la décision du voyageur, certains amènent sur les sites que tout le monde va voir (le Mont Saint Michel), où il y aura beaucoup d'autres visiteurs. D'autres conduisent sur des sites auxquels seuls les initiés s'intéressent (le pavillon de Flaubert). Choisir le chemin correspond à organiser la navigation, avec plus ou moins d'alternatives : paramétrer un GPS avec son domicile comme point de départ et point de retour sécurise l'itinéraire ; se référer aux panneaux suscite une part d'incertitude dans le voyage. Pour tracer un chemin, certains outils équipent individuellement les voyageurs (GPS), alors que d'autres s'appuient sur des ressources communes (panneaux). L'ordre et l'accès aux sites retenus par le visiteur s'élaborent grâce à des prises issues de différents dispositifs : les étoiles et avis pour établir un ordre, le GPS et panneaux pour mettre en œuvre l'accès. Ces artefacts techniques se complètent et se concurrencent pour produire l'itinéraire final.

Cette métaphore du voyage sert à étudier l'exploration du web par un internaute en l'envisageant comme l'exploration d'une région par un voyageur. L'internaute doit identifier les sites qu'il veut visiter et parcourir le chemin y conduisant. De même que le voyageur collecte des informations sur les lieux à voir et s'équipe pour le voyage, l'internaute se saisit de signaux affichés par les dispositifs pour visiter les pages de la toile. Quels sont les artefacts numériques qui permettent de découvrir le web, de choisir des pages et des chemins ?

a) L'idéologie collaborative pour hiérarchiser les documents numériques

Parmi la multitude de pages sur des multitudes de sites, quelle page mérite un intérêt ? La surabondance d'informations en ligne naît de l'abaissement des barrières à la production de contenus mais aussi du fait que les méta-informations qui contribuaient hors ligne à estimer la valeur des contenus ont perdu de leur consistance. Pour évaluer son intérêt pour un document, l'internaute ne peut plus se référer aux prises traditionnelles comme l'auteur et l'éditeur, qui sont devenus pluriels ou inconnus ; il est submergé de bien d'autres méta-informations hétérogènes, comme le compteur de pages vues ou la moyenne des notes attribuées par ceux qui se sont investis en jury.

Ces méta-informations ne sont pas disponibles immédiatement pour l'internaute qui navigue en ligne : elles doivent être alimentées par des utilisateurs ou producteurs du web et présentées dans les dispositifs. La hiérarchisation des documents numériques est produite *ex ante* par les services à partir d'algorithmes, l'ordre des documents est élaboré au cours de la navigation de l'internaute en fonction de son attention aux différents critères de hiérarchisation et de sa curiosité.

Dès les premières friches du web, l'idéal de la mise en réseau qui inspire l'architecture technique s'est imposé comme référence pour les dispositifs de classement du web.

L'hypothèse de « la sagesse des foules » (Surowiecki, 2008) s'illustre à travers différents exemples dans les champs de la psychologie, la sociologie ou l'économie. Elle propose qu'un collectif d'individus est à même de prendre une décision plus optimale que celle que prendrait individuellement chacun des individus qui composent le collectif, à condition que le groupe formé respecte trois caractéristiques : (1) laisser s'exprimer les avis de manière indépendante les uns des autres ; (2) intégrer une diversité des points de vue et expériences, ce qui nécessite que le collectif se compose d'un nombre assez grand d'individus ; (3) agréger les avis de manière décentralisée et extérieure au collectif.

Les instigateurs du web ont conceptualisé une architecture technique du réseau correspondant en plusieurs points à ce fonctionnement. Le réseau commuté se déploie techniquement en créant des nœuds d'activité multiples, qui décentralisent l'intelligence et produisent globalement le réseau des réseaux. La hiérarchisation des contenus se calque sur ce fonctionnement technique : aucun internaute ne sait individuellement évaluer la qualité d'un contenu autrement qu'à l'étalon de son intérêt personnel ; mais en agrégeant les mesures d'un collectif d'internautes diversifiés, le classement résultant est optimal. Pour évaluer la qualité d'un document numérique, les dispositifs ont implémenté techniquement les principes de l'intelligence collective en agrégeant les « avis » des « jury ». Reste à identifier les modalités d'un avis et les individus qui forment le jury.

Dans les exemples de Surowiecki, la combinaison d'une foule et d'une expression produit une modalité de la sagesse : soit la connaissance, qui s'appuie sur l'expérience individuelle du plus grand nombre ; soit la coordination, qui nécessite de penser le système global pour mettre en avant les expériences singulières de chacun ; soit la coopération, qui réunit des acteurs concurrents pour produire une résultante non-concurrente. Dominique Cardon identifie lui trois principes d'ordre du web : l'éditorialisation, où des « documentalistes » rangent les documents numériques suivant un classement singulier mais accrédité par leur statut ; l'autorité, où la citation par les pairs correspond à une reconnaissance méritocratique, à travers l'utilisation des citations entre sites ; l'audience, où l'activité du plus grand nombre atteste de l'intérêt d'un contenu (Cardon, 2011)¹.

En reprenant l'histoire des dispositifs du web, trois générations d'outils permettant d'évaluer la valeur des documents peuvent être identifiées. À chaque étape, les dispositifs de l'ordre produisent un des modèles identifiés par Dominique Cardon, en s'appuyant sur un des modes de fonctionnement décrit par Surowiecki. Les dispositifs ont ainsi à charge de fixer le périmètre des internautes « jury » considérés, et les modalités d'expression de leur « avis ». La résultante est un ordre donné du web, une « sagesse » correspondant aux techniques et aux usages en cours en fonction du stade de développement et de diffusion du réseau.

¹ Dominique Cardon identifie ensuite l'émergence d'un ordre affinitaire et d'un ordre lié à la fraîcheur des contenus, qui interviendront dans la deuxième partie de ce chapitre.

La sélection des lecteurs-auteurs dans les 'filter blogs'

Le dispositif primitif permettant de hiérarchiser collectivement des pages web consiste à créer une ou des pages signalant la sélection de chacun à tous : ces ancêtres des blogs se regroupent sous l'appellation des « *filter blogs* ». Un *filter blog* est un site web produit par un internaute, qui signale à la communauté les sites qu'il a appréciés et qui actualise cette sélection au fil de ses explorations. Cette publication crée une « navigation sociale » (Dieberger, 1997), un internaute allant de pages de contenus en pages de blogueurs.

Dans ses travaux sur l'émergence du format des blogs, Ignacio Siles a collecté plusieurs témoignages sur le sens et la pratique de cette activité. Les *filter blogs* ont émergé de la frustration des pionniers de ne pouvoir tout découvrir. Ces précurseurs se sont alors lancés dans la « distillation » du web en « préfiltrant » chacun les contenus (Siles, 2011). Le principe d'ordonnement du web mobilisé ici est celui de l'éditorialisation : les internautes sélectionnent les pages en fonction de leurs compétences propres. L'innovation technique qui permet de mettre en œuvre cette coordination pour éditer l'ordre du web est la simplification de la publication : chaque internaute peut publier son *best-of* du web sans qu'il n'y ait au préalable de sélection de son expression. La sagesse de cette sélection résulte de la coordination des pionniers du web, auteurs-lecteurs qui naviguent mus par leur curiosité respective et restituent dans le même temps le produit filtré par leurs compétences. La sélection individuelle prend sens si la pratique est assurée par une population de pionniers assez conséquente pour distiller le web avec des intérêts variés.

Les *filter blogs* se stabilisent grâce à la standardisation des outils utilisés et l'identification d'une communauté de contributeurs. Côté outil, les plates-formes d'édition telles que Blogger ont fixé des cadres de pratique, par exemple l'affichage des publications par ordre ante-chronologique, ou les *blogrolls* pour signaler les sites similaires. Dès 1998, certains sites comme le *filter blog* Memepool¹ sont devenus collaboratifs en permettant à plusieurs internautes de les alimenter. Côté communauté, auteurs masculins et contenus technophiles distinguent les *filter blogs* des journaux intimes des *diaristes*, écrits par des femmes à partir de contenus personnels. Une étude récente montre qu'aujourd'hui, les contributeurs des *filter blogs* ont un profil particulier : il s'agit d'hommes avec des pratiques médiatiques singulières, rejetant notamment la télévision (Wei, 2009). Pour ces sélectionneurs du web, l'ambition n'est pas uniquement de filtrer le web de manière collaborative, mais aussi de se différencier des médias *mainstream*. Cette spécialisation de la pratique réduit le web exploré et crée une présélection des sites suivant un critère idéologique plutôt qu'une coordination *ex ante*.

Après l'étape de stabilisation, les dispositifs d'éditorialisation du web par tous ont donc perduré dans une communauté de spécialistes. La « sagesse » ne peut plus être atteinte

¹ <http://www.memepool.com/>, créé en 1998, consulté le 03/11/2013, dernier post daté du 03/05/2013

dès lors que les profils formant le collectif sont singuliers. La sélection produite par les *filter blogs* n'est donc plus suffisamment hétéroclite pour que tout internaute trouve dans cette production une hiérarchisation du web opérante pour son exploration. D'autant que les internautes qui accèdent au web ne sont plus tous impliqués dans la production collective : les auteurs-lecteurs deviennent marginaux parmi les utilisateurs. L'ordre du web proposé par les pionniers à travers les *filter blogs* est devenu un ordre singulier ou idéologique. La critique qui s'esquisse dans cette trajectoire des *filter blogs* est celle d'un découpage du web : au lieu d'envisager l'ensemble des contenus possibles, l'éditorialisation coordonnée s'est au fil de temps concentrée sur certains contenus pour n'explorer qu'un espace restreint du web.

Les citations des producteurs dans le PageRank

Le fait que tout internaute puisse écrire et publier sur le web a conduit au développement d'une deuxième génération de hiérarchisation du web basée sur un classement par les auteurs. Le classement proposé est implémenté par le moteur de recherche Google en 1998, dans son algorithme nommé *PageRank*, avant que le principe ne soit repris par les autres moteurs de recherche. Un moteur de recherche met en regard deux univers : celui des contenus publics et intemporels du web, et celui des besoins particuliers et ponctuels d'un internaute. L'enjeu est donc d'abord de pouvoir indexer l'ensemble des contenus disponibles en ligne, et ensuite de répondre à une requête particulière d'un internaute en restituant les contenus de manière hiérarchisée par rapport à la demande. La performance du moteur de recherche résulte à la fois de l'indexation exhaustive du réseau et de la pertinence des contenus renvoyés à chaque requête. Les premiers moteurs sémantiques n'étaient pas parvenus à construire des réponses satisfaisantes aux recherches des internautes en se fondant sur le registre lexical des contenus. En 1998, le *PageRank* de Google révolutionne le mode d'évaluation d'une page en utilisant les références à cette page dans les autres pages pour calculer un indice de qualité du contenu. L'algorithme se base donc sur les liens pointant vers une page pour estimer la qualité de cette dernière, la citation conférant au contenu un caractère d'autorité. Ce principe applique le modèle d'évaluation des citations académiques aux pages web en considérant le lien hypertexte comme matérialisation de la citation (Cardon, 2013).

Le *PageRank* comme principe de hiérarchisation du web met en œuvre un mécanisme collectif, puisque dans ce service ce n'est pas une citation qui détermine la valeur d'un document, mais l'ensemble des citations. Référencer l'ensemble des documents et l'ensemble des citations présente une certaine complexité, les technologies de *crawling* et d'algorithmique sont donc nécessaires pour afficher les résultats hiérarchisés à une requête. Grâce aux technologies, le *PageRank* fait coopérer, au sens de Surowiecki, les internautes-auteurs concurrents, pour produire un ordre des documents non concurrent. Par rapport à l'éditorialisation, l'autorité implémentée s'appuie sur l'établissement des

internauts-auteurs, professionnels ou amateurs qui écrivent, comme individus compétents pour attribuer à un contenu une valeur. Les citations doivent rester naturelles et « oublier » qu'elles génèrent un ordre¹, afin de maintenir les conditions de diversité et d'indépendance nécessaires pour produire un classement optimal des documents.

En revanche, cet ordre du web ne crée pas une intelligence de la navigation, puisque le moteur ne montre pas aux internautes les références utilisées pour justifier l'intérêt d'un contenu². Le moteur de recherche décharge l'internaute de la responsabilité de sa navigation en affichant des documents dans un ordre non explicité. Deux artefacts techniques de Google témoignent que le moteur de recherche remonte dans la gestion de contenus, et déborde de la production d'un ordre pour s'immiscer dans la navigation de l'internaute.

Depuis 2005, le moteur de recherche applique une personnalisation des résultats, par exemple en fonction de l'historique de recherche de l'internaute et de sa localisation. La personnalisation « nettoie » les résultats de recherche en intégrant le contexte de la requête. Par exemple, la personnalisation supprime les contenus en thaïlandais en réponse à une requête en français. Ce fonctionnement élabore implicitement un contexte à la recherche pour filtrer les documents présentés. Il se base sur une connaissance partielle de l'internaute, qui ne retient que ce que l'utilisateur livre dans son historique de navigation ou dans le paramétrage de son ordinateur. Cette connaissance partielle de l'utilisateur et la difficulté à construire des indicateurs d'opinion des contenus alimentent le doute sur la légitimité de la hiérarchisation filtrée pour l'internaute par rapport à la hiérarchisation produite par les références. Pour continuer la comparaison avec les voyages, la personnalisation des résultats de Google revient à restreindre la liste des sites à visiter sur la base des précédents voyages de l'internaute. Le dispositif est donc critiquable suivant l'argument qu'il « découpe » le web en espaces autonomes préalablement à la requête de l'internaute, restreignant l'accès des internautes à certains sites.

La deuxième fonctionnalité interférant avec l'ordre méritocratique créé par Google est le mécanisme de suggestion de requêtes. La formulation d'une requête dans un moteur de recherche est critique, puisqu'en elle réside la singularité de l'internaute, et qu'elle conditionne la qualité des premiers résultats répondants à la question. Or depuis l'été 2008, Google propose des « suggestions » de requête à partir des premiers caractères saisis par l'internaute, sur la base des requêtes des autres internautes et de l'historique

¹ La qualité des résultats de recherche est détériorée si les sites web optimisent leurs contenus pour les rendre visibles par le moteur, ou si par exemple des fermes de contenus génèrent des « fausses » citations.

² Alors que dans les citations académiques, les références sont indiquées pour qu'un chercheur puisse remonter la bibliographie et ainsi acquérir à son tour des compétences pour produire des contenus.

des recherches de l'internaute. Cette fonction est appelée « saisie semi-automatique ». Par exemple, il suffit de commencer à taper « serge brin la... » pour que Google propose « *sergueï brin larry page* » en corrigeant au passage le prénom de Brin. Google justifie cette fonctionnalité comme un avantage pour « *gagner du temps et identifier les erreurs* »¹. La saisie semi-automatique des requêtes ne peut pas être désactivée car elle rend la requête « *plus rapide et plus facile* ». Cet artefact technique implémente un pan de sagesse des internautes qui font des recherches sur le moteur : si un internaute a besoin d'une information, probablement d'autres ont au préalable déjà cherché cette même information. Mais ce petit morceau de code risque de standardiser un langage des requêtes en ligne et par là les documents consultés². Cette fonctionnalité amènerait donc un « aplanissement » du web, une réduction de l'immensité à un plan unique en oubliant le web plus profond. Les suggestions de requête dans Google amènent tous les visiteurs au Mont-Saint-Michel, alors que certains auraient privilégié le Mont des Cats.

Le hit-parade des lecteurs dans Digg

Avec la généralisation des connexions haut-débit et la démocratisation du numérique, les internautes deviennent encore plus nombreux et les internautes-lecteurs deviennent massivement plus importants que les internautes-auteurs. La troisième génération des services de hiérarchisation du web va donc chercher à capter la contribution des internautes-lecteurs pour ordonner les contenus du web : *Digg* fait voter les internautes pour construire sa Une, Google Actu s'appuie sur la mesure d'audience des articles pour construire un panorama de presse.

Ces deux services s'inscrivent toujours dans les conditions permettant de mettre en œuvre une sagesse des foules, la foule étant cette fois-ci l'ensemble des internautes-lecteurs, et la modalité d'expression étant le clic signifiant l'intérêt pour un contenu. La connaissance est le modèle de contribution utilisée par les compteurs, puisqu'ils permettent de savoir ce sur quoi un grand nombre d'internautes clique. L'ordre de l'audience est ainsi mis en œuvre par ces dispositifs, en utilisant la généralisation de l'accès au web et la production de compteurs.

Digg a été le pionnier des sites proposant aux internautes de cliquer pour hiérarchiser l'information. Sur la plate-forme lancée en 2004, tout inscrit peut poster un lien d'information qu'il estime intéressant et les votes (les « *digg* ») des autres internautes sur ce lien le feront progressivement remonter en Une sur la page d'accueil. La Une de Digg classe donc l'information sur le web par le vote des internautes-lecteurs. Ce service a

¹ <https://support.google.com/websearch/answer/106230?hl=fr>, consulté le 03/03/2014

² Le détournement de la saisie semi-automatique par les pratiques de *Google bombing* montre le versant négatif du mécanisme : en envoyant automatiquement de nombreuses recherches dans le moteur, les hackers peuvent rendre ces recherches suggérées et ainsi orienter l'intérêt des internautes sur les formulations construites par les *bombers*.

remplacé le « tous journalistes » proposé par les sites comme Agoravox par un « tous éditeurs », soulignant le fait que c'est l'ordonnancement des informations plutôt que leur production qui devient déterminant dans un contexte de surabondance informationnelle. Précurseur et très soutenu par les pionniers du web¹, Digg a fait l'objet à la fois d'expérimentations et de critiques qui montrent les aléas de ce classement.

Au lancement de Digg, tous les votes comptaient à égalité, comme dans les modalités mises en œuvre par les élections directes. Mais dès mai 2005, Digg développe un algorithme non linéaire pour pondérer les votes plébiscitant un sujet : un poids est attribué à chaque électeur, en fonction du volume de ses contributions antérieures au global et par rubrique². Les experts d'une rubrique ont alors un poids plus important que les novices dans la construction de l'ordre de cette rubrique. Ce fonctionnement permet de reconnaître la légitimité de certains contributeurs. Mais le fait que certains internautes aient un poids plus important que d'autres a amené des stratégies d'influence et créé un marché. Un des contributeurs les plus actifs reconnaît en 2006 avoir été rémunéré pour promouvoir certains sujets et, inversement, des lobbies ont financé des campagnes de dénigrement de contenu (puisque le vote « négatif » existe sur Digg). On constate ainsi qu'un algorithme tenant compte du profil et de l'activité des votants est nécessaire pour valoriser l'expertise propre à chaque contributeur, mais que cet algorithme ouvre des inégalités et fausse la sélection.

Quand bien même l'ordre produit par les votes resterait efficient, la pratique mécaniquement valorisée par Digg est celle du militantisme. Force est de constater que les contenus à la Une de la plate-forme ne sont ni les contenus à la Une des médias, ni les contenus les plus lus : il s'agit en général d'un mélange d'articles engagés pour une cause³. Au lieu d'observer une démocratisation de l'expression à tous les internautes-lecteurs, Digg témoigne de l'investissement du web par les médiactivistes décrits par Dominique Cardon et Fabien Granjon (Cardon, Granjon, 2011). Les internautes qui proposent des contenus se mobilisent pour une cause et sont actifs dans des communautés, donc proches des auteurs-lecteurs du web. Les électeurs sur la plate-forme votent pour des contenus initialement engagés, l'information résultante est donc nécessairement orientée. Aucun service s'inspirant du fonctionnement de Digg⁴ n'a réussi à susciter les

¹ Kevin Rose est un des quatre fondateurs de Digg, et avait auparavant une notoriété dans les médias. Les fondateurs de eBay et Netscape ont tous deux investi dans la première levée de fonds de Digg.

² L'algorithme de Digg n'a jamais été publié, la pondération envisagée résulte donc de l'analyse des controverses sur le sujet.

³ Quand il ne s'agit pas de contenus illégaux : en 2007, Digg a dû renoncer à supprimer de sa plate-forme le code de décryptage des DRM de DVD et Blue-Ray, n'ayant pas les moyens de supprimer ce contenu perpétuellement republié par la communauté.

⁴ De nombreux services se sont lancés sur le principe de Digg en spécialisant le champ ou raffinant le mécanisme : Scoopeo, Wikio, etc.

votes d'internautes non-actifs. L'expression des internautes-lecteurs à partir de leur vote¹ ne produit donc pas un ordre collaboratif des documents numériques, mais un ordre engagé.

Est-ce qu'utiliser le clic de la navigation comme indicateur de l'intérêt peut résoudre ce biais sur les internautes électeurs et permettre de prendre en compte l'expression de tous les internautes ? Le clic aurait pour avantage de s'exercer sur des contenus plus larges que ceux sélectionnés par des militants et de ne pas demander d'engagement supplémentaire à l'internaute. Cette mesure donnerait une voix à la majorité silencieuse. L'idée de mesurer l'audience d'un contenu pour juger de sa qualité a fait l'objet de nombreux débats à l'occasion du lancement de Google News. Ce service, développé par la firme de Mountain View, affiche un panorama de presse au fil de la journée, en hiérarchisant les contenus collectés sur les médias suivant un algorithme intégrant l'audience du contenu et sa fraîcheur². Cette construction de l'actualité n'est toutefois pas assez documentée pour que l'on puisse intégrer ce cas dans une étude de l'ordre en ligne. Mais retenons la critique selon laquelle le fait de hiérarchiser des informations à partir de leur audience conduit à une standardisation des contenus : cet indicateur valorise les documents universels et socialement neutres, ceux qui réunissent tous les publics.

Si cet indicateur est pris en compte dans la hiérarchisation du web, il contribue à homogénéiser l'information à laquelle on accède³. En rendant signifiant dans une mesure globale un clic protéiforme, qui résulte autant d'un intérêt civique que d'une curiosité ponctuelle ou de l'oisiveté de l'internaute, le web crée une mesure qui aplanit le web. Cette mesure alimente de plus le mécanisme que les industries culturelles désignent comme la prime aux leaders : un contenu lu va apparaître en tête du hit-parade et être ainsi encore plus lu. L'internaute n'exerce pas son « vote » dans une liste de contenus engagés comme sur Digg, mais dans une liste de contenus homogénéisés. On retrouve ici

¹ On peut souligner que les principes démocratiques d'égalité des votes mis en œuvre par Digg sont différents des principes délibératifs et autogouvernés mis en œuvre dans d'autres espaces du web, comme Wikipédia (Auray, 2009).

² L'algorithme de classement des actualités restant secret, le rôle de l'audience n'est pas exactement connu. Le fonctionnement de Google News fait l'objet de critique de la part des éditeurs, qui considèrent le service comme du pillage de leur contenu et des revenus publicitaires de leur Une. Cette question socio-économique ne sera pas abordée ici, voir Rebillard, Smyrniotis, 2011 au sujet des infomédiaires.

³ Newswhip a sorti des unes de journaux en fonction des sujets plébiscités par l'audience et Slate l'a imité en France (<http://blog.newswhip.com/index.php/2014/03/people-powered-front-pages-rock>; <http://www.slate.fr/france/84557/partage-internautes-vs-une-quotidiens-nationaux>; consultés le 02/04/2014). Comparer systématiquement la une éditoriale d'un journal et la une de l'audience du journal serait bien sûr une étude intéressante, mais ne peut être mise en œuvre sans le concours des médias pour accéder à l'audience des articles. La distinction entre la consultation et les pratiques de partage d'information en ligne sera abordée au chapitre 2.

la critique de l'« aplanissement » du web à un espace documentaire commun à tous mais réduit au plus petit intérêt commun.

En synthèse, retracer l'histoire des services présentant une hiérarchisation des contenus du web rend compte des artefacts qui ont outillés l'exploration des internautes pour choisir les pages à visiter. Les premiers dispositifs (*filter blogs*) proposaient aux internautes de classer collectivement le web avec un principe d'édition, c'est-à-dire en coordonnant la sélection des contenus. Ces services ont été élaborés par la communauté des pionniers du web, auteurs-lecteurs. Le web continuant d'augmenter et les internautes n'étant plus toujours auteurs, les algorithmes de recherche ont élaboré un mécanisme de coopération des internautes-auteurs pour classer le web suivant l'autorité attribuée par les citations. Cet ordre idéal est limité soit par la personnalisation réduisant l'exploration des individus, soit par l'homogénéisation des recherches amenant les internautes sur certains contenus. Enfin, les services se basant sur les votes et les clics ont le mérite d'avoir l'ambition d'intégrer l'activité de tous les internautes-lecteurs dans la hiérarchisation du web, en se basant sur la connaissance des activités de tous. Mais l'ordre par l'audience aboutit à une mise en exergue des sujets engagés quand l'activité est observée par le vote des internautes, ou à des sujets neutres quand l'activité est observée par la consultation des internautes.

L'ordre des contenus mis en œuvre par les dispositifs techniques contribue donc soit (1) à découper le web, l'internaute ne voyant que ce que le dispositif lui propose de manière personnalisée ; soit (2) à aplanir le web, l'internaute étant amené sur les contenus universels consultés par tous. On retrouve ici les travers de la sagesse des foules, le risque de la perte d'autonomie des individus dans un groupe et la difficulté de production d'une résultante même en tirant profit des capacités de calcul et d'agrégation des calculateurs externes et autonomes.

b) S'équiper et équiper le web pour accéder aux documents numériques

Reprenons le problème de départ. « Tout » est accessible dans le monde numérique ... mais à quoi accède-t-on réellement? Une fois passé un sentiment initial de vertige, légitime, le premier clic permet de relativiser cette illusion du tout : le web ne peut être que partiellement exploré. Le chemin, au sens de la navigation, devient alors déterminant de ce à quoi l'on accède. Des services proposent de trouver en ligne la réponse à une requête, à partir d'une intention initiale ; l'internaute explore alors le web dirigé par cette intention. Mais le web permet aussi une navigation dite « par sérendipité » : de liens en liens, l'internaute accède à des contenus qu'il n'avait pas nécessairement l'intention de consulter mais qui suscitent son intérêt. En travaillant la sérendipité avec le concept d'« exploration curieuse » dans les communautés de hackers, Nicolas Auray montre que cette forme de navigation est un équilibre ténu entre activité et passivité, entre conduire et se laisser conduire (Auray, 2009). La sérendipité n'est pas qu'une rêverie, mais bien une activité en clic. L'exploration du web ne répond pas uniquement à une recherche utilitariste, au prisme de laquelle seule l'évaluation par rapport à la requête compte. Elle résulte d'une expérience dynamique rendue possible et entretenue par la navigation.

L'activité de l'internaute est donc productrice du chemin d'accès aux contenus, et ce chemin se décrit avec trois caractéristiques. Chaque accès se réalise parmi une *pluralité*

de chemins possibles. On peut accéder à une page web après une requête dans un moteur de recherche ou suite à la consultation de multiples *posts* sur un forum conseillant ce contenu ; après avoir lu un article à charge contre l'auteur de cette page dans un journal ou convaincu par les éloges d'un billet dans un blog. Du fait de cette pluralité des parcours, le chemin d'accès à un contenu numérique est aussi *indéterminé*. Il n'est pas possible de savoir *a priori* d'où vient un internaute quand il lit un contenu, ni même où il ira après. L'internaute peut suivre le lien « page suivante » comme il peut prendre une autre porte de sortie. Ces éléments contribuent à rendre la navigation *singulière* à chaque internaute. Chacun construit son parcours : aux vues de la taille du web et de la multiplicité des chemins possibles, la probabilité que deux internautes consultent les mêmes pages par le même chemin est quasi nulle. En ligne, les documents numériques sont accessibles à tous, mais l'accès est propre à chacun.

Ces caractéristiques de l'accès aux contenus numériques renforcent le rôle de la navigation comme prise de la réception d'un contenu. De même que la sérendipité n'est pas propre au web mais s'accorde particulièrement à l'espace virtuel, le rôle de l'expérience et de l'accès dans la réception d'un contenu était déjà présent dans la réception des médias traditionnels¹ mais s'accroît avec les pratiques numériques. On ne regarde pas une série de la même manière en laissant la télévision allumée après le journal télévisé, ou en ayant attendu toute la semaine le nouvel épisode. Les travaux sur les biens culturels montrent aussi le rôle de la jaquette d'un CD ou de la première de couverture d'un roman dans l'expérience esthétique du mélomane et du lecteur (Hennion, 2007). De multiples artefacts techniques, informationnels et esthétiques viennent contribuer à la réception d'un document numérique en plus des prises qui aiguillent l'accès à un contenu hors ligne. Et parmi ces prises, le chemin d'accès au contenu est le résultat de l'activité de l'internaute, qui investit son parcours et en est responsable. On « n'aime » pas un page web de la même manière suivant qu'elle s'est affichée en tête de Google en un clic et deux demi secondes, ou qu'elle apparaît en point d'orgue d'une longue navigation alternant recherche ciblée et curiosité. La valeur accordée à un document numérique porte la marque de la trajectoire qui a conduit l'internaute à ce point, et cette marque est d'autant plus fortement appliquée au contenu que l'internaute l'a lui-même produite.

L'implication cognitive de l'internaute dans sa navigation le met en capacité de se saisir d'un lien plutôt que de celui d'à côté, en mobilisant ses ressources et compétences personnelles. Ainsi, l'internaute accède à un contenu « entraîné » par son activité. « Entraîné » au sens de la transitivité du mouvement, l'internaute clique parce que le lien est présent sur la page et qu'il va voir plus loin, dans une dynamique. « Entraîné » comme

¹ Cette exploration curieuse est aussi possible hors ligne, par exemple en flânant dans une bibliothèque (voir le discours de Umberto Eco à Milan le 10 mars 1981, <http://www.samuelhuet.com/linguistique/47-folio/556-umberto-eco-de-bibliotheca.html> consulté le 03/04/2015). La dématérialisation des contenus rend l'espace de la flânerie plus large, et les prises pour se saisir d'un contenu plus hétéroclites.

un sportif, l'exploration de l'internaute résultant d'une pratique antérieure et régulière de la navigation. On peut alors distinguer deux types d'outils conçus pour entraîner l'internaute dans son exploration. Les premiers sont ceux qui permettent à l'internaute de se préparer par une pratique régulière ; l'internaute s'équipe pour sa navigation, comme un voyageur achèterait une carte sur laquelle il peut se repérer et écrire ses commentaires. Les deuxièmes sont ceux qui déposent des informations sur les contenus pour entraîner le mouvement de l'exploration ; il s'agit de mailler l'espace des documents numériques de chemins d'accès cognitifs, comme le voyageur se référerait à des panneaux et cairns construits par les services ou les autres randonneurs plutôt qu'à une carte personnelle. Un troisième type d'outils, plus récent, développe ces deux approches dans les pratiques expertes, relevant des compétences acquises au sujet d'un centre d'intérêt ou d'une passion plutôt que sur l'expérience du web. Comme pour la hiérarchisation des documents, les artefacts d'accès aux contenus numériques ont progressivement intégré des fonctions pour permettre à des internautes de plus en plus nombreux et non technophiles d'utiliser les outils. Mais l'histoire qui va être retracé ici aboutit à l'échec de la démocratisation de ces outils, comme s'ils passaient à côté de la généralisation de l'accès au web pour devenir au fil du temps des outils de spécialistes.

Des bookmarks à Netvibes : personnaliser l'entrée dans le web

Comment démarrer une navigation ? Le point de départ de l'exploration est déterminant, puisqu'il place l'internaute à un point de l'espace numérique qui lui permettra d'accéder à certains contenus en plus ou moins de clics. Et c'est bien sur cette ligne de *start* que se sont concentrés les premiers dispositifs techniques équipant l'exploration. Des *bookmarks* aux portails, les services du web que l'internaute paramètre pour démarrer sa navigation s'attellent à gérer la schizophrénie de l'exploration : retenir les contenus vus et partir en voir d'autres, s'ancrer aux documents connus pour découvrir des espaces inconnus.

Le premier dispositif qui rentre dans cette catégorie d'artefacts de la navigation est un véritable pilier historique : il s'agit du *bookmark*, ou marque-page. Mettre une page web en *bookmark* consiste à poser un signet sur cette page pour y revenir directement en un clic, l'url étant mémorisée dans le navigateur lui-même. Cette fonctionnalité est développée par l'aïeul des navigateurs, Mozaic, en 1993, sur la base d'une spécification technique explicite : être un élément persistant d'une connexion à l'autre.

Le problème des internautes de l'époque est en effet que fermer son navigateur revient à effacer toutes les pages parcourues. Entre le papier que l'on peut archiver et la télévision ou la radio que l'on ne peut que suivre, le web construit un intermédiaire où des éléments du flux de la navigation peuvent être facilement retenus.

Les *bookmarks* s'attachent au plus petit niveau de granularité des documents numériques, l'URL. On peut enregistrer en marque page aussi bien un blog qu'un billet de blog, un site de média qu'un article du site, la page d'accueil d'un service qu'une requête pré-paramétrée sur le service. Ils pointent précisément un contenu consulté à un moment donné, mais l'artefact technique ne gère pas de temporalité : si un internaute a mis un signet sur un blog spécialisé, il devra retourner sur ce blog pour voir s'il y a un nouveau billet ; et il prend le risque que le blog s'arrête et que le contenu ne soit plus actualisé,

mémorisant ainsi des contenus virtuellement morts. Le *bookmark* est ainsi un dispositif qui permet de retenir un contenu numérique pour pouvoir y retourner, sans gérer l'enrichissement du web. Il fige la sélection, et de ce fait est plus approprié à la mémorisation de sources éditoriales ou de services qu'à la sélection de contenus ou de connaissances. Un *bookmark* est plus approprié sur le monde.fr que sur un article du monde.fr, probablement périmé à la prochaine consultation.

L'internaute risque donc, en partant des *bookmarks*, de répéter sa navigation, sans accéder aux nouveautés publiées. Pour pallier au manque d'actualisation des marques-pages, les concepteurs du web ont fixé en 2003 le standard RSS (*Really Simple Syndication*). Tout site intégrant ce standard dans son développement fait automatiquement figurer ses nouvelles publications dans un flux, dit flux RSS. La standardisation du format est fixée pour que tout internaute puisse paramétrer ses flux RSS personnels en ciblant les sources et les rubriques ou mots clés qui l'intéressent ; les dispositifs dits lecteur de flux RSS permettent de suivre les nouveaux contenus issus de ces flux. Ce fonctionnement est toutefois resté trop complexe pour passer dans les usages grand public¹. À défaut d'un succès d'usage, les flux RSS ont à l'époque gagné un succès technique et sont prisés par les services de « portails ». Ceux-ci agrègent des flux RSS pour présenter des actualités et services à l'internaute. Ils foisonnent de 1997 à 2005, proposant chacun une porte d'entrée dans le *world wide web*. Des médias aux pages personnelles, chaque site intègre à cette époque sur sa *homepage* des liens vers les quatre coins du web, pour couvrir le plus de chemins numériques possibles et tenter de convaincre le plus d'internautes de mettre cette page en page d'accueil du navigateur².

Netvibes, lancé en 2005, incarne le service de portail personnalisé et dynamique le plus abouti grâce à deux éléments : un très riche annuaire de connecteurs à des services et des flux RSS, et une personnalisation simple et attrayante pour le grand public grâce à l'implémentation du *drag and drop*. Le service propose ainsi à tout internaute de créer son portail personnel en découpant sa page d'accueil en rubriques. Chaque rubrique intègre un flux de contenus actualisés grâce aux RSS ou des services connectés, par exemple les champs de recherche pour avoir l'horaire d'un train. Le *drag and drop* permet de dimensionner et déplacer les rubriques sur la page simplement avec la souris, ce qui rend le portail d'autant plus personnel : non seulement l'internaute choisit les contenus, mais en plus il maquette sa page suivant une organisation de l'espace de l'écran qui lui est propre. Netvibes a connu un certain succès, sans que le nombre d'utilisateurs n'ait jamais été confirmé, mais annonçant une rentabilité fin 2009³. Le rachat

¹ Les sites éditeurs comme les médias ont pré-paramétré des flux pour que les internautes s'y abonnent, reproduisant les modèles de diffusion des médias hors ligne. Mais l'arrêt du lecteur Google RSS, en 2013, témoigne de l'extinction de l'usage des flux RSS par les internautes.

² Stratégie qui assure un nombre de visiteurs important et permet d'augmenter les revenus publicitaires de la page.

³ <http://blog.netvibes.com/fr/netvibes-atteint-la-rentabilite/>, consulté le 10/05/2015.

de la société en 2012 atteste toutefois du fait que le portail personnalisé tombe en désuétude dans les usages courants, alors que les *dashboard* et outils du web, développés en parallèle, se placent sur un marché BtoB plus monétisable. Cette construction des portails a été reprise par les fournisseurs d'accès Internet : pour beaucoup d'internautes, leur page de démarrage est le portail de leur opérateur, un mix d'actualités, de mails, de météo, de programme télé, et de moteur de recherche, mêlant donc contenus et services en ligne. En bleu, rouge, ou orange en fonction de l'opérateur, le navigateur de chaque internaute est plus ou moins personnalisé suivant le niveau d'informations configurées par l'utilisateur. L'élaboration d'une page d'accueil personnalisée est ainsi facilitée pour le grand public non technophile en restreignant les fonctionnalités permettant de choisir l'affichage ou les sources de contenus.

Les *bookmarks* forment pour l'internaute un réservoir de sa navigation, permettant de retrouver des contenus déjà consultés ; la version dynamique de cette fonction, les portails, amène directement au point de départ que sont les sources présélectionnées. L'internaute met en place ces repères pour commencer sa navigation à partir des éditeurs et contenus préalablement évalués. Mais deux critiques peuvent être formulées concernant ces outils d'entrée dans le web. Tout d'abord, les usages de ces dispositifs se sont peu développés avec la démocratisation du web. Les *bookmarks* sont préférentiellement remplacés par la « mosaïque » des navigateurs, qui affiche au lancement les neuf sites les plus fréquentés par l'internaute¹. Ni les flux RSS ni Netvibes n'ont réussi à se généraliser. Le portail construit par l'internaute est remplacé par le portail de son opérateur, intégrant automatiquement des informations personnelles et une sélection générique de service. L'internaute grand public privilégie donc les fonctions automatiques ou pré-paramétrées pour rentrer sur le web, plutôt que de s'équiper personnellement avec des outils techniques pour gérer sa navigation. Ensuite, ces outils incitent à la répétition d'un itinéraire plutôt qu'à une « exploration curieuse » : l'univers du web consulté est celui préalablement référencé. Les *bookmarks* entretiennent des pratiques rituelles en figeant les points de départ. Les portails créent un flux d'informations issues de certaines sources comme dans les médias traditionnels, sans susciter une découverte des « choses » hétérogènes du web.

Pour retrouver une exploration ouverte, il faut alors retourner la démarche, en équipant non pas l'internaute mais l'internet lui-même.

¹ Depuis les années 2012, Firefox puis Chrome et Internet Explorer ont instauré dans leur page de lancement une « mosaïque » de sites, mosaïque personnalisée en fonction de la navigation de l'utilisateur de l'ordinateur. Il est en fait possible de les paramétrer, c'est-à-dire de fixer les propositions de certains sites ou de forcer d'autres propositions. Je ne connais pas d'étude sur l'usage des *bookmarks*, ni de la personnalisation des navigateurs.

La catégorisation et les tags : déposer des traces pour baliser la navigation

Après s'être intéressé au point de départ d'une navigation, il convient donc de s'intéresser au web dans son ensemble. Pour aborder ce vaste espace, des dispositifs ont cherché à « ranger » le web. Cette entreprise a initialement été menée par des annuaires, comme Yahoo!

L'idée était de construire une structure exhaustive et hiérarchique de rubriques permettant d'accéder aux sites web en réduisant, rubrique après rubrique, la liste des sites du champ déterminé. L'annuaire ouvert DMOZ¹ poursuit cette entreprise : il affiche aujourd'hui 17 catégories à la racine, chaque catégorie se découpe en sous-catégories, le nombre de sous-catégories est potentiellement illimité, et le nombre d'embranchements pour accéder aux « feuilles » que sont les sites n'est pas préalablement connu. Un internaute peut cheminer dans 5 sous-catégories avant de se voir proposer la liste de sites répondant à sa navigation, comme il peut s'enfoncer dans 15 sous-catégories sans toutefois aboutir à une sélection de sites à son goût. L'ambition de structurer le web en rubriques et de proposer un classement éditorial universel à l'ensemble de la production est devenue une tâche infinie et inefficace pour la navigation. Et surtout, c'est une pratique centralisée, que le web a délégué à chaque internaute pour produire un rangement individuel et ouvert plutôt que unique et totalisant.

Ranger le web participe à « l'entraînement » de l'internaute. Se saisir et s'approprier un contenu au fil des clics et des lectures incarne la part active de la réception. Or en classant un document dans une ou plusieurs catégorie(s), l'individu formalise sa réception et rebondit sur les contenus suivants. Ainsi des dispositifs proposent aux internautes de classer le web en « taguant » les pages web de leur navigation. Taguer consiste à déposer un signet sur le contenu, sans se l'approprier matériellement en l'enregistrant dans son ordinateur. Le principe est de mailler sa navigation au fil des pages en y déposant des balises, lesquelles serviront autant à retrouver les pages visitées qu'à s'orienter dans l'espace numérique pour aller consulter d'autres documents. Comme le petit poucet s'aventure dans la forêt parce qu'il a marqué le chemin effectué, les signets en ligne marquent les pages visitées pour sédimenter le savoir et soutenir la suite de l'exploration.

L'usage des tags pour catégoriser les pages web s'est développé dans le courant des années 2000, notamment avec le service Delicious. Développé dans la continuité des *filter blogs* par le créateur de Memepool, Delicious consiste en une extension de navigateur pour faire figurer un bouton « taguer » dans le menu. En cliquant sur ce bouton alors qu'il est sur une page web, l'internaute ouvre une fenêtre *ad-hoc* lui demandant de renseigner les tags qu'il veut affecter à la page en cours de consultation. Ces tags sont en texte libre (sans référence à une liste préétablie, mais avec une suggestion des tags déjà utilisés) et un internaute peut renseigner plusieurs tags pour une même page. La catégorisation est

¹ <http://www.dmoz.fr/>, consulté le 25/04/2015.

ainsi personnelle, multiple et évolutive. Les pages taguées sont archivées et accessibles sur le compte de l'utilisateur, avec des fonctions de tri par tag, combinaison de tags, date, etc. Le service Delicious est ainsi né en proposant à l'internaute de construire son fond documentaire de contenus structurés, mais en laissant ce fond en ligne. Il simplifie la sélection des contenus par rapport aux *filter blogs* et l'enrichit des catégories.

En étudiant les pratiques de taguage de photos dans Flickr, Maxime Crepel montre que cette activité répond initialement à un besoin individuel et peut déboucher sur une activité discursive et interactionnelle (Crepel, 2011). Par défaut, les pages des utilisateurs de Flickr sont publiques, ce qui permet à un internaute de suivre un compte par exemple d'un expert, ou de visualiser les catégories formulées par d'autres utilisateurs sur une page quand lui-même la tague. Maxime Crepel montre alors que, sur Flickr, les tags sont le support d'une réception sociale : pour un internaute, voir les mots apposés sur un contenu numérique permet de questionner sa réception et de renégocier son point de vue. Les outils de discussion en ligne (commentaires, messages privés, forums, etc.) poussent à élaborer des catégories à plusieurs et à débattre de l'activité. Par exemple, les utilisateurs de Flickr entreprennent des « concours photo » pour structurer et développer une catégorie.

Sur Delicious, les communautés de contributeurs se forment autour de centres d'intérêt communs. Ceux qui utilisent les mêmes tags ou taguant les mêmes contenus se reconnaissent, s'évaluent à travers leur sélection et la solidité des catégories créées. Cette communauté « discute » numériquement en de la catégorisation, sans qu'un consensus soit nécessaire pour rendre universel le rangement puisque les tags d'un contenu peuvent être multiples. Cette discussion soutient la réception d'une photo et entraîne pour la suite de la navigation. Par exemple, elle peut permettre d'identifier qu'une catégorie de photos est sous-alimentée et inciter une navigation dédiée à identifier les contenus liés à ce tag. Ainsi, des interactions se créent, virtuellement, autour du choix des tags et de la cohérence des catégories créées. Le taguage ouvre donc une sociabilité dans une pratique initialement individuelle. Internet permet de rendre public un rangement privé, activant la part collective et sociale de la réception. Le rangement du web délégué aux internautes n'est pas réduit à l'échelle de chaque individu mais augmenté grâce à la mise en réseau des internautes.

La catégorisation permet alors à un niveau collaboratif d'indexer le web. Les tags en ligne produisent des « *folksonomies* » : il s'agit des taxonomies produites par les foules, une indexation permettant de référencer les contenus à partir de la multi-indexation attribuée par une communauté d'amateurs. Collectivement, les tags qualifient potentiellement davantage de documents numériques qu'une indexation uniquement gérée par des professionnels ne le ferait, permettant ainsi de nourrir de méta-

informations plus de pages web¹. La taxonomie produite est aussi potentiellement différente d'une catégorisation universelle pensée en chambre par des experts du domaine ou de la veille, ou d'une catégorisation personnelle renseignée par l'auteur du document et indiquée dans les méta-informations. Les *folksonomies* produites par les publics se distinguent de l'indexation des documentalistes et des descriptions des auteurs, et d'après l'hypothèse de l'intelligence des foules elles seraient optimales. Cette utopie de la production collective du savoir reste toutefois à étudier empiriquement. La critique latente souligne le risque inhérent à la diversité des tags : la description en mots est à la fois le support de l'interaction et une ouverture à la multiplicité des rangements ; mais la multiplicité des tags utilisés produit une catégorisation floue, impossible à synthétiser en une connaissance commune et opérante pour asseoir l'accès aux contenus.

Dans notre histoire des dispositifs d'accès aux contenus, Delicious distille l'ordre impulsé par le filtrage du web en simplifiant l'activité individuelle de sélection. A ce titre, le service ne produit pas de l'ordre, puisque les contenus ne sont pas évalués mais rangés. Le rangement est un support à la navigation : en permettant de mémoriser des contenus avec la méta-information de leur catégorisation, le service fait un pas de plus vers l'appropriation par rapport aux *bookmarks*. Delicious participe de plus à une re-documentarisation du web grâce aux catégorisations individuelles multiples, alimentant une pratique de rangement perpétuel essentielle dans la construction du savoir. Ranger, déranger, re-ranger les contenus construit les savoirs individuellement (Weinberger, 2007) ; en ligne ces pratiques deviennent collectives et renouvelle la catégorisation universelle des annuaires.

Mais Delicious est resté un outil confidentiel : de même que les contributeurs des *filter blogs* sont des pionniers du web, les utilisateurs de tags sur Flickr sont des « pro-am »² de la photo, et les utilisateurs de Delicious sont finalement des « pro-am » de la veille documentaire. La singularité des « tagueurs » n'est pas un frein à la production d'un rangement du web collectif, mais ce rangement des contenus du web n'est pas utilisé par les internautes qui n'y participent pas. Ces derniers n'entreprennent pas une lecture sociale des contenus en observant les tags apposés par d'autres, ni ne naviguent *via* les comptes Delicious d'experts identifiés. Le rangement du web par les « pro-am » de la veille en ligne ne sert pas de voies d'accès aux documents numériques pour le grand public. De plus, la lecture du web avec les multiples catégories des tagueurs génèrent une navigation en pointillé : les tags sont éparpillés sur le web et il n'est pas possible de

¹ Il faudrait toutefois vérifier empiriquement que les tags ne se concentrent pas sur les contenus phare du web, comme on observe une concentration de l'audience des contenus audiovisuels en replay (Beuscart et al. 2012), une concentration des commentaires / like sur certains articles de presse (Bastard, 2013), etc.

² ProAm pour « Professionnels - amateurs » : terme qui qualifie les individus qui, sans devenir professionnels d'une activité, ont développé une pratique quasi professionnelle grâce aux outils du numérique. Ceux-ci permettent de mesurer la pratique, publiciser l'activité, trouver une communauté, etc. Voir Flichy, 2010.

chaîner les contenus s'y rattachant dans un ordre de progression et d'exploration. Les balises en ligne ne donnent pas le chemin intermédiaire entre deux contenus, ils obligent donc l'internaute à naviguer avec des fragments plutôt qu'avec un cheminement. Après l'abandon des *bookmarks* et la désuétude de Netvibes, Delicious illustre donc un nouvel échec dans l'histoire des outils d'accès aux contenus web.

Scoop.it : explorer les chemins d'accès par l'expertise

Repartir des pratiques professionnelles permet alors de rendre compte des difficultés inhérentes au rangement du web et de mieux comprendre les contraintes rencontrées par les services précédemment présentés. Dans les travaux des sciences de l'information et de la documentation, la notion de « curation » s'applique à ces nouvelles pratiques de rangement du web¹. Une définition de cette activité a été proposée par Véronique Mesguich : « *la curation consiste à repérer divers contenus numériques en fonction d'une thématique donnée, de sélectionner et filtrer les plus pertinents, de les organiser, les structurer à travers un dispositif de scénographie et d'en favoriser la diffusion.* » (Mesguich, 2012 : 25). La curation a drapé les pratiques traditionnelles des documentalistes dans les artefacts du web, et a été érigée en fondement d'une nouvelle ère de production de la connaissance². Son potentiel comme modalité d'accès aux documents numériques réside dans l'utopie de la généralisation de cette pratique semi-professionnelle : si tout internaute peut entreprendre d'être curateur du web, l'espace documentaire géré serait plus vaste et le rangement plus juste.

L'exemple de Scoop.it peut servir de cas d'étude puisque ce service s'adresse explicitement aux curateurs. Les fondateurs entendent résoudre le problème de « l'infobésité » en réunissant sur leur plate-forme les veilleurs qui assurent chacun un rangement des contenus web suivant leurs champs d'expertise. Le curateur crée son ou ses *topics*, ou centres d'intérêt, et les alimente en « scoopant » des contenus au cours de sa navigation, c'est-à-dire en taguant ces contenus pour les intégrer dans ses *topics*. Les liens retenus par les curateurs de Scoop.it s'accumulent dans les pages de *topics* du curateur. Le service éditorialise les contenus sélectionnés avec les titre, visuel, source et résumé des contenus, si l'auteur initial a renseigné ces éléments. Il permet d'ajouter un *insight* du curateur, par exemple un commentaire personnel. Scoop.it améliore ainsi les *tags* de Delicious par une mise en page des contenus enrichis de leurs méta-informations et de leur évaluation.

Scoop.it insiste sur l'individualisation de la pratique de rangement, en outillant l'espace des curateurs. La page présentant les *topics* d'un utilisateur peut être affichée, et chaque *topic* peut être personnalisées par le curateur. Cette visualisation donne l'impression que

¹ Voir le numéro spécial de la revue *Documentalistes*, Mesguich et al. 2012.

² La curation a fait l'objet d'une volumineuse production de guides, manuels, conférences et formations. Voir par exemple Rosenbaum, 2011.

les contenus sont empaquetés par le curateur. Scoop.it singularise le rangement sans produire un rangement collaboratif. C'est la qualification d'expert qui est démocratisée par Scoop.it plus que l'accès aux contenus. Pour un internaute qui explore le web, il est en effet possible de se référer au *topic* d'un curateur pour accéder à des contenus éditorialisés et catégorisés. Alors que Delicious alternait la navigation de la page d'un contenu à la page d'un individu, les topics d'un expert servent de point central pour explorer un champ donné. Les curateurs deviennent les nouveaux intermédiaires d'accès aux contenus. Le passage d'un web des documents à un web des individus apparaît ici, puisqu'un internaute ne suit plus une édition mais un profil. La critique de la curation prend prétexte de cette mise en page personnalisée et de la médiation du curateur : pour certains, cette activité pille les contenus en les sortant du contexte de leur production initiale et en volant à l'auteur son public¹.

La collaboration autour du rangement du web est en fait mise en œuvre à un autre niveau que la publicisation : Scoop.it alimente le curateur avec les sources qu'il a lui-même pré-renseignées, les comptes Twitter qu'il suit, et l'analyse textuelle des contenus *scoopés*. La technologie revient à l'analyse sémantique, inefficace au début des moteurs de recherche, mais ici aiguisée par son application à des contenus et des profils préfiltrés. Elle élargit le *sourcing* des experts d'un domaine, à partir d'un corpus de documents sur lequel le moteur de Scoop.it peut apprendre les centres d'intérêts de l'internaute. Au lieu de rendre visible les tags du web pour tout internaute, le service balise l'exploration des curateurs en réalisant une analyse textuelle des contenus tagués par la communauté des curateurs, et c'est au curateur de servir de guide aux internautes non experts.

Le salut de l'exploration du web viendrait ainsi de l'élargissement des experts à une large population ayant des centres d'intérêt et passions divers. La semi-professionnalisation des communautés, par des outils comme Scoop.it, est un développement visant à garantir l'élargissement des contenus sélectionnés. Véronique Mesguich rappelle toutefois que les outils proposés produisent un impact indéterminé sur l'exploration : est-ce que la curation ouvre la navigation à des espaces élargis ou la restreint dans des communautés homophiles ? Le risque est d'assécher la diversité des sources et de réduire l'exploration des contenus à une portion du web en fonction des espaces couverts par les experts, mais en laissant vierges de toute re-documentarisation des espaces non investis par des passionnés.

Retenons le fait que l'expertise serait le levier d'une certaine curiosité pour résoudre l'entrée dans le web et le balisage des contenus pour la communauté. La passion équipe l'internaute de compétences personnelles et lui permet de s'aventurer dans des

¹ Juridiquement, l'affichage d'une vignette et d'une citation courte est en effet soumis au droit d'auteur s'il n'est pas réalisé par un journaliste (Battisti, 2012). Ce qui voudrait dire que les curateurs doivent demander l'accord de l'auteur pour les citer... pratique qui n'est bien évidemment pas mise en œuvre en ligne.

territoires en ligne sans revenir toujours aux mêmes contenus ou aux mêmes sources. Avec le risque que ce mode d'exploration du web ne soit entrepris que par des individus spécifiques et n'ouvre des voies d'accès aux contenus que sur certains champs. Dans la comparaison avec les voyages, l'expertise serait par exemple une compétence de géologue ou de botaniste, qui permet à un randonneur d'explorer une région sans avoir besoin de ses cartes ou des panneaux, son repérage technique de l'environnement étant suffisant pour créer un itinéraire et rendre cet itinéraire accessible à des non-experts. L'expertise est l'équipement « infra » de l'internaute, au sens où il est permanent et mobilisable pour la navigation en ligne, ainsi que pour d'autres activités.

Pour conclure sur l'histoire des dispositifs équipant l'internaute pour accéder aux documents numériques, rappelons que les artefacts du début du web s'attachaient à proposer un point de départ à la navigation de l'internaute par une mémorisation des contenus retenus et sources sélectionnées. Cet amorçage de la navigation conduit toutefois l'internaute à répéter des trajets connus plutôt qu'à découvrir l'hétérogénéité du web. L'autre équipement, incarné par les *tags*, permet de baliser individuellement et collectivement la navigation, sans s'approprier les documents mais en les rangeant et en discutant leur réception. Cet éparpillement d'étiquettes crée toutefois un chemin en pointillé, où l'internaute sait identifier tel et tel contenu mais ne peut pas élaborer un chemin progressif à son exploration. La principale critique que l'on peut formuler sur ces outils précurseurs de la curation reste leur faible diffusion : ces équipements personnels et publics sont en effet utilisés par des spécialistes de la veille documentaire ou des experts, sans que leur intermédiation ne serve d'aiguillage dans la navigation des internautes. Le grand public ne se saisit pas des outils de l'exploration numérique à moins d'être un amateur, et il ne sait pas se référer aux chemins et prises créés par les amateurs car celles-ci restent disséminées.

c) L'espace aplani ou découpé, des accès répétés ou en pointillé

Les outils qui accompagnent les deux activités de l'internaute dans son exploration du web des documents, les outils de l'ordre et les outils de l'accès, doivent répondre au problème de l'exploration du web rencontré par chaque internaute. Le tableau 1 présente les services étudiés dans une perspective historique, pour montrer la sédimentation des artefacts techniques et l'enrichissement progressif des outils par la technique.

L'ordre des documents numériques est implémenté par les dispositifs en s'appuyant sur une légitimité éditoriale (par les pionniers, producteurs-lecteurs), une autorité experte (par les internautes-auteurs) ou une validation par le volume d'audience (par les internautes-lecteurs). Ces dispositifs mettent en exergue la double nature des informations : bien universel et partagé socialement ; bien personnel consommé et approprié individuellement. Cette double nature amène les dispositifs à aplanir le web sur un plan unique, réduisant un contenu à sa fonction de vecteur social, objet commun à discuter ; ou à découper le web en sous-espaces personnalisés, où les contenus sont évalués au seul prisme d'un intérêt personnel.

Tableau 1 : *Il y a « trop de choses » sur le web*

Activité	Hiérarchiser les contenus (destination - ordre)	Accéder aux contenus (chemin - accès)
1995-2000	Filter blogs Ordre éditorial - Coordination Pionniers	Bookmark S'équiper Grand public
2000-2005	Google PageRank Ordre autorité - Coopération Internaute-auteurs	Delicious, Tags Equiper le web Veilleur
2005-2010	Digg Ordre audience - Connaissance Internaute-lecteurs	Scoop.it S'approprier le web Expert
Tension	Aplanir le web / découper le web	Itinéraire répété / itinéraire en pointillé

Dans la navigation individuelle, les modalités d'accès aux documents s'inscrivent dans des pratiques de gestion documentaire où la catégorisation, chemin d'accès cognitif du document, est déterminante. Elles entraînent l'internaute en lui faisant pratiquer un accès au contenu et en initiant la dynamique de navigation. Les dispositifs tentent d'ouvrir une exploration du web à partir des expertises, sans toutefois permettre au grand public de se saisir de la catégorisation des experts pour accéder aux contenus. Les dispositifs de navigation orientent l'internaute en facilitant la répétition de trajectoires pré-validées à partir d'un point de départ, ou en éparpillant des étiquettes sur des contenus en pointillés, sans accompagner le chemin entre deux contenus.

Finalement, les artefacts de l'ordre et de l'accès s'empilent et s'entremêlent dans la multitude de méta-informations des documents numériques et ne résolvent pas tout à fait le problème du « trop de choses ». Les outils de hiérarchisation de l'information se heurtent à un volume de contenus encore et toujours plus important, et ne savent plus gérer un indicateur global et unique pour évaluer la valeur d'un contenu. Les outils d'accès aux contenus sont utilisés par des semi-professionnels de la veille et de l'information, mais n'offrent pas au grand public du web démocratisé un moyen de s'orienter en ligne sans se référer à une programmation ou à une navigation ritualisée.