

Fluidification informationnelle et intermédiaires sociotechniques

Cette communication a pour objectif de mettre en rapport les pratiques actuelles des jeunes en matière de recherche d'information sur le web avec les avancées techniques en cours dans le domaine. Derrière la fluidité apparente de la recherche d'information sur le web s'agglomèrent des technologies dont la complexité va croissant. L'écart se creuse entre les connaissances nécessaires à leur compréhension et leur usage.

Consommation plurimédias

L'entrée des micro-ordinateurs dans les foyers et leur connexion systématique à internet a recentré les activités de communication et de consommation des médias autour de nouveaux objets. L'incitation à rester en permanence « connectés » avec les mobiles, la messagerie instantanée ou le web touche prioritairement ceux que l'on appelle les « digital natives¹ » : les jeunes férus de technologies de communication dont l'apprentissage informatique a tôt commencé.

Dans ces environnements, les institutions de diffusion (comme la presse, la radio et la télévision) se sont vus concurrencés pendant leur migration en ligne par les « nouveaux médias », dont les contenus sont souvent produits par les internautes eux-mêmes.

La consommation des médias s'individualise au sein des familles qui ne se rassemblent plus toujours autour du téléviseur et se recentre autour de l'utilisateur qui « choisit », parmi un éventail infini de possibilités, des items de plus en plus fins (les grains d'informations).

On considère que la recherche d'informations sert essentiellement à documenter trois activités principales : le jeu (ex : recherche de procédures d'installations, de solutions, de découverte de jeux), la consommation de médias (ex : les fans vont retrouver sur Youtube les montages concernant leurs favoris), le travail scolaire (préparer un exposé oral ou écrit).

Il s'agit d'une activité contextuelle dont la pratique individuelle ou en collectif est dépendante d'un environnement sous contraintes (chambre, salon, médiathèque, équipement scolaire encadré par les professeurs ou accessible en libre-service).

L'intensité de la recherche est variable selon les besoins, dépendante du zapping et fonction de la mobilité.

Les intermédiaires de l'information

La recherche d'informations fait appel à deux sortes d'intermédiaires. Les intermédiaires humains qui s'inscrivent dans une chaîne critique de validation de l'information sont les parents, documentalistes, professeurs, conseillers d'orientation, etc. Ils sont susceptibles d'apporter une aide mais aussi de porter des jugements critiques, de sélectionner les informations pertinentes en fonction d'une discussion qu'ils auront initiée.

Les intermédiaires sociotechniques sont les automates dont se servent les « digital natives » pour accéder directement à l'information, à savoir principalement les moteurs de recherche.

L'offre des intermédiaires sociotechniques a pour caractéristique de fournir des résultats en nombre quasi-illimité, dont la pertinence est toute relative tout en bénéficiant de la « neutralité » apparente de la technique : un moteur de recherche ne porte pas de jugements et évite ainsi toute confrontation critique. En effet, les intermédiaires sociotechniques ne

¹ L'emploi de cette appellation controversée (<http://www.digitalnative.org/#about>) vise à définir un public auquel j'enseigne en IUT (bac+1+2+3) dans une filière technologique.

préjugent pas (encore) de ce que vous allez faire des informations.

Les « digital natives » prennent ainsi progressivement l'habitude d'utiliser des automates pour rechercher des informations : ils activent leurs réseaux sociaux (ex : Myspace), se servent des liens de la blogosphère (ex : Skyrock) ou accèdent aux ressources par le tagging (ex : Youtube, Dailymotion, Flickr, etc.) tout en publiant eux-mêmes leur production (ex : Deviantart).

Bien souvent, l'accès facile à des « mines d'informations » les pousse à ignorer le savoir-faire des infomédiaires (intermédiaires humains dont c'est le métier).

Cette profusion d'informations conjuguée au zapping permanent entre médias engendre des problèmes que l'on peut rapporter aux modes cognitifs identifiés par Katherine Hayles². Elle différencie la « Deep attention » ou capacité de se concentrer sur un seul objet, en ignorant les stimuli extérieurs pendant une longue période, de l'« Hyper attention » ou capacité de réaliser plusieurs tâches en parallèle en changeant d'objectif, avec une faible tolérance à l'ennui. Ainsi les « digital natives » appartiendraient à la seconde catégorie et souffriraient d'un déficit d'attention dans les tâches qui requièrent de la concentration, ce qui signifie des difficultés à apprendre, réfléchir, écrire.

Face au déluge informationnel et à la multiplication des sources (fiables ou non), un certain nombre de difficultés sont observables quant à l'évaluation de l'information. Pour la majorité des jeunes étudiants, les résultats fournis par les moteurs de recherche sont acceptés sans aucune distance critique.

La « neutralité » apparente des moteurs est entretenue par un vocabulaire choisi par l'industrie liée au marketing web : par exemple, le référencement créé par l'indexation des moteurs est prétendu « organique » ou « naturel » par les agences de communication spécialisées dans la création et la production web. On constate ainsi l'insuffisance de la culture informationnelle des jeunes. Le fonctionnement technique et commercial des moteurs de recherche et systèmes de recommandations n'est pas connu et les processus de médiatisation ainsi que la place des infomédiaires dans la fabrique des contenus comme les relations publiques, le journalisme ou le marketing politique sont totalement ignorés.

A cet état de fait, il faut ajouter la réalité de la fluidification des informations par les médias. L'information journalistique file, court, passe et se trouve prise dans un flux sans fin. Lorsque l'information est erronée, seule la contrainte juridique obtient sa rectification publique qui se produit en second plan et finit par apparaître accessoire, voire inexistante au public pressé. De plus, les éventuelles « suites » données aux « nouvelles » sont rares et superficielles. La conclusion que certains en tirent est qu'il importe peu que les informations soient erronées, l'essentiel étant de passer à autre chose.

Du côté des « digital natives », si on adopte un point de vue pessimiste, on peut en déduire que l'importance de la véracité de l'information décroît, ce qui entraîne des comportements de copier/coller inadéquats, réalisés sans réelle réflexion ; l'essentiel étant de remplir (copions ce que nous trouvons...). Inutile de creuser, passons à autre chose (et ne cherchons pas à comprendre).

Quel dénouement rechercher alors ? Une solution serait d'initier à la fabrique de l'information en apprenant à choisir, à croiser les sources, en se posant les questions de ce qu'est une information vraie, une source fiable. Autant de préludes à l'apprentissage de l'évaluation et à la prévention du plagiat. Le dépassement du copier/coller est rendu possible

² <http://media08.wordpress.com/2008/01/17/my-article-on-hyper-and-deep-attention/>

par l'apprentissage ludique de la réécriture qui doit être enseignée avec patience et méthode. Comment former de futurs auteurs si l'on prétend encore qu'ils doivent tout sortir d'eux-mêmes alors qu'ils surnagent avec peine dans l'océan informationnel ? La publication « personnelle » devrait être enseignée.

Comment ? En manipulant les outils de publication et de recherche qui sont utilisés actuellement sur le web et en s'insérant de manière intelligente dans les réseaux sociaux. Pour ce qui est de la fabrique collective de l'information, nous avons les wikis. La maîtrise d'un concept éditorial peut être approchée par la tenue d'un blog (que dire et ne pas dire...). La recherche thématique d'informations peut être pratiquée par l'intermédiaire d'abonnements à des fils RSS spécialisés.

Enfin, un retour critique encadré sur les produits réalisés est indispensable.

Réalités techniques du web

Nous assistons à une complexification croissante qui concerne l'afflux des sources d'information, l'hétérogénéité des données, les calculs et les traitements potentiels sur ces données.

Si un web documentaire s'est affirmé, repérable par ses métadonnées interopérables avec les standards de description XML et RDF, un « nouveau » web est en préparation qui risque d'ajouter une couche d'opacité à celle existante des moteurs de recherche.

Le web « actuel » a connu deux phases.

La première phase que l'on peut situer entre 1996 et 2004 a vu l'introduction d'outils générant des sites web « clés-en-mains » : blogs, wikis, cms. Ceci a démocratisé l'accès à la publication personnelle sur le web, auparavant réservée aux techniciens du html. Les plateformes de publication en ligne (blogs, wikis) ont aussi facilité et stimulé la collaboration massivement distribuée dont les principales caractéristiques sont la production de nouveaux formats éditoriaux (microblogging) et une forte connectivité (blogosphère).

A cette première période, correspond la recherche d'informations par moteurs. Les moteurs de recherche indexent les documents sur le web et les classent en fonction de calculs statistiques. Ces documents sont vus par les algorithmes comme des jeux de mots-clés auxquels on attribue des poids et sur lesquels on effectue des statistiques.

Les principales caractéristiques sont un classement des réponses aux usagers basé sur la fréquence des mots-clés des pages web ET en fonction d'une analyse des liens « entrants » qui détermine le référencement des sites web (qui pointe vers qui ?).

La deuxième phase du web actuel qui prend son essor en 2004 jusqu'à aujourd'hui est propulsée par l'industrie logicielle qui crée le « web2.0 » avec le principal objectif de provoquer la migration « en ligne » d'activités pratiquées sur ordinateur personnel. Un certain nombre d'innovations émaillent cette période : messagerie électronique (gmail) ; logiciels de traitement de texte, de retouches d'images, de montage, de gestion du temps se pratiquent sur des serveurs distants ; le partage de ressources par tagging est rendu possible et les usagers indexent avec leurs propres mots images, vidéos, textes et signets. Pour cela, les internautes s'identifient avec leur adresse électronique, ce qui permet l'enregistrement des traces d'usages liées à leur profil.

Cependant, la principale conséquence est que les internautes ne se contentent plus de consulter les ressources mais apprennent à manipuler des « ressources informationnelles » et entrent ainsi dans un cycle de création-consommation-critique. Les services liés au réseautage

se sont multipliés et la création de la valeur par les activités des internautes (UGC³) est devenue une manne gratuite pour les expérimentateurs de nouveaux services.

Du point de vue de la recherche d'informations, cette deuxième phase est constituée par l'ajout d'une couche de filtrage aux résultats fournis par les moteurs de recherche : l'évaluation par les usagers repérable par leurs tags et leurs appréciations. Un processus d'indexation collaborative des documents sur le web est amorcé avec la généralisation du « tagging », l'évaluation et le classement possible des ressources par les usagers (avec la généralisation des commentaires et votes).

Les principales conséquences sont l'appropriation par les usagers de la notion de métadonnées avec les « folksonomies ». L'utilisation de métadonnées interopérables se démocratise (fils RSS encodés en RDF).

Un enjeu serait de faire se rejoindre les pratiques des usagers avec celles des professionnels (tags et thésaurus par exemple) dans le cas de corpus délimités.

Promesses du web sémantique

En 2001, Tim Berners Lee du W3C annonçait « un vaste espace d'échanges de ressources entre machines permettant l'exploitation de grands volumes d'informations et de services variés, aidant les utilisateurs en les libérant d'une bonne partie de leur travail de recherche et de combinaison de ces ressources »⁴. L'objectif affiché était d'informer les gens ET les machines en rendant le web « intelligent ».

Le web sémantique prépare un changement de paradigme. L'objectif principal du web actuel est de transmettre des contenus intelligibles dont l'observable restitué sur écran doit être fidèle à ce que veulent les auteurs. Pour le « web sémantique », l'objectif principal qui s'ajoute au précédent est d'améliorer la pertinence des contenus en opérant des traitements sur eux. Comment ? En préjugant ce que veulent les internautes par des formalisations qui produisent de l'intelligibilité.

On mesure combien cette ambition pourrait faire craindre l'abandon des questionnements concernant l'évaluation de la recherche d'informations. Puisque la complexité accrue qui s'annonce sera difficilement compréhensible aux « opérateurs humains ».

Dans le web sémantique appelé aussi « web de données », les ressources ne sont plus seulement des documents mais aussi des programmes, des interfaces web, des services qui sont caractérisés par leurs métadonnées⁵ (voir Fig. 1).

Chaque ressource est reliée aux autres par des relations hiérarchiques ou associatives (étiquetage des liens), indiquant leur appartenance à des domaines dont les connaissances sont formalisées (définition et relation entre les termes) et organisées en ontologies.

Un exemple concret est la différenciation faite par le W3C⁶ entre URI, URL et URN⁷. Alors que l'URL identifie une ressource par son adresse, l'URI ajoute à celle-ci une identification

³ User generated content

⁴ Citation extraite de « Une introduction au web sémantique ». Journée Action Spécifique Web Sémantique du 07/05/2003. Philippe Laublet.
<http://www.lalic.paris4.sorbonne.fr/stic/mai2003/programme070503.html>

⁵ <http://www.w3.org/2001/12/semweb-fin/w3csw>

⁶ Semantic Web Activity, Koivunen and Miller, 2001

⁷ Uri : uniform resource identifier

Url : uniform resource locator

Urn : uniform resource name

par son nom (URN). Ce détachement de la localisation de la ressource dans sa description la rend libre pour d'autres usages. Elle n'est plus subordonnée à l'accès et au référencement par les moteurs car elle existe par son nom dans la mémoire documentaire. Un méta-classement se met donc lentement en place.

En adoptant le tagging, les internautes sont entrés de plain-pied dans le web sémantique sans en être spécifiquement conscients. Alors que les liens entre documentation et informatique sont depuis longtemps réalisés, la modélisation de l'accès aux ressources du web entame une étape supplémentaire qui s'avère laborieuse et complexe. La question est alors posée de comment associer à leur tour les enseignements et de continuer à former à la critique ?

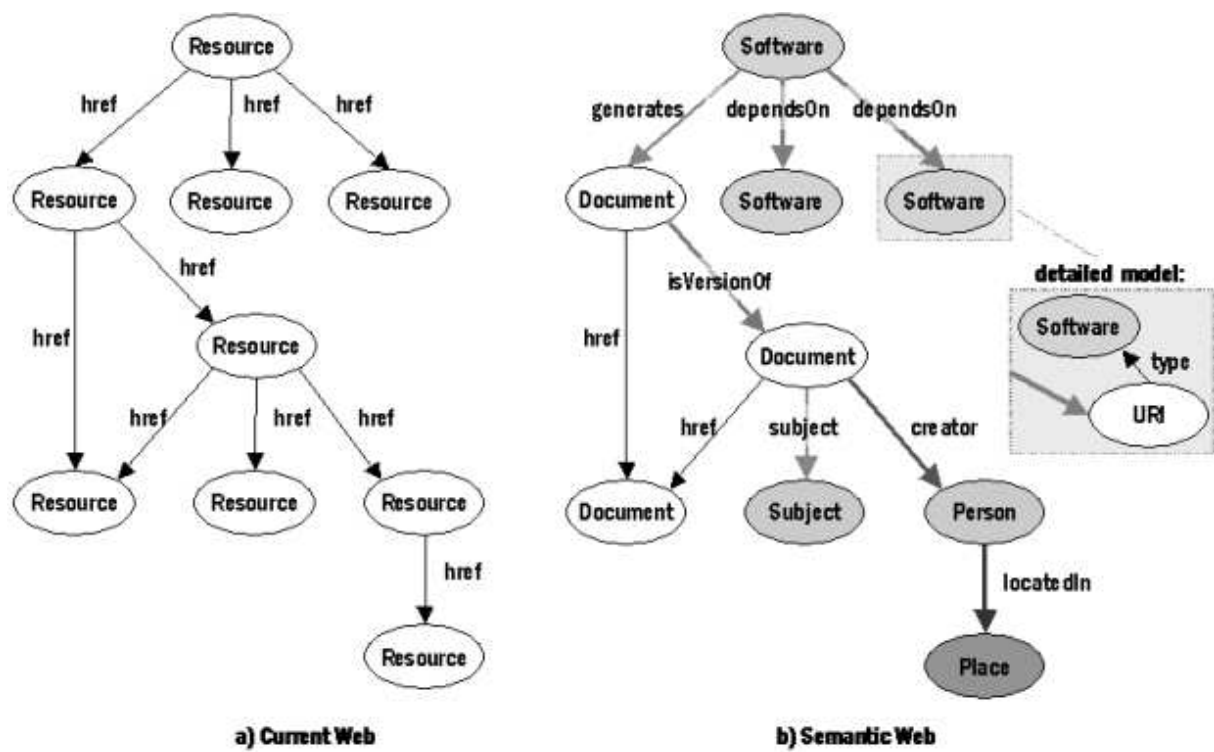


Figure 1 - Représentation de l'activité dans le web sémantique : liens et ressources sont typés (Koivunen and Miller, 2001).