

Des bons mots au bon document. Comment éduquer à l'usage des mots-clés efficaces pour accéder à la pertinence documentaire.

Olivier Le Deuff
Cersic-Erellif, Rennes 2.
oledeuff@gmail.com

Résumé : La diversité des types de discours utilisés sur Internet entraîne une confusion chez l'élève qui ne parvient pas toujours à optimiser sa recherche. La formation à la maîtrise et à la production de l'information permet l'acquisition d'habiletés essentielles.

Mots-clés : document, recherche documentaire, mots-clés, information literacy, élève, documentaliste, négligences.

Abstract : The diversity of speeches used on Internet involves a confusion for the pupil who always does not succeed in optimizing his research. Information literacy increase documentary skills.

Keywords : document, information literacy, keywords, student, negligences.

Intoduction.

Les stratégies de recherche des élèves sur Internet suscitent parfois interrogations et débats. L'apprentissage des mots-clés s'avère difficile car il s'agit pour l'élève de réfléchir et de catégoriser afin de convertir des idées en mots-clés. Ce processus n'est pas naturel et implique des capacités d'abstraction qui font souvent défaut d'autant plus que sur Internet nous sommes confrontés à des types de discours fort variés selon les sites visités et les applications utilisées. Les élèves rencontrent des difficultés pour effectuer la distinction entre tous ces discours et ne parviennent pas toujours à savoir quel langage employer au bon moment à bon escient. Ces confusions expliquent également les négligences et autres erreurs de lecture commises fréquemment.

Ces difficultés sont à rapprocher des représentations erronées de l'Internet qu'ont souvent les élèves. Notre propos vise à démontrer que la capacité à utiliser des mots-clés pour effectuer une recherche s'acquiert progressivement et que la production de contenus sur Internet par l'élève augmente ses capacités notamment grâce aux possibilités offertes par le Web 2.0. Nos observations et expériences ont été principalement établies dans un collège du sud Manche avec des élèves ayant entre 10 et 16 ans et montrent la nécessité d'une réelle formation à la recherche et la production d'informations (information literacy). Nos travaux cherchent à étudier les relations entre le document et les usagers et les problèmes communicationnels qui demeurent en dépit des avancées des traitements automatisés.

1 .La médiation des moteurs.

La recherche de documents évolue et n'est pas uniquement axée sur les moteurs de recherche. En effet de nouvelles stratégies documentaires voient le jour opérant une « redocumentarisation » comme le nomme le dernier document du RTP-Doc (RTP-Doc 2006). Ainsi ce « nouvel âge de la navigation » inclut les possibilités de s'abonner à des flux d'informations notamment. Malgré tout les moteurs de recherche demeurent très nettement le moyen le plus fréquemment utilisé par les élèves pour rechercher de l'information. Les bases de données sont peu usitées par ces derniers qui demeurent bien souvent aux portes du Web invisible mais bon nombre d'enseignants n'échappent pas

non plus à la règle. De fait les difficultés pour accéder à la pertinence documentaire¹ sont relativement élevées dans ce cadre. La recherche documentaire sur Internet recèle quelques complexités et l'usage des élèves tend au contraire vers une simplicité proche du guichet unique de l'information. Nous songeons évidemment à Google mais certains élèves tapent parfois leur recherche dans la barre d'adresses ce qui les renvoie à la page par défaut dans Internet Explorer.

L'usage des moteurs étant désormais le moyen utilisé par les plus jeunes pour accéder à l'information, les documents papiers sont souvent évités voire ignorés (Six Degrés 2006). De même la tentation de « foncer » sans réflexion est omniprésente. Les documentalistes de collège peuvent en témoigner. Il faut sans cesse rappeler l'importance d'une réflexion préalable. La domination du moteur Google se trouve d'autant plus forte que les notions abstraites de l'Internet sont peu maîtrisées. Nous avons pu constater cet état de fait dans la formation et l'évaluation dans le cadre du B2I (Brevet informatique et internet). La confusion des discours commence avant même que l'utilisateur ne soit connecté. Il s'avère que pour beaucoup il existe une forte confusion entre les mots « navigateur », « moteur de recherche », « logiciel », « fournisseur d'accès ». Mais il est clair qu'un seul mot ressort de ce désordre sémantique : « Google ». Après plusieurs séances de travail sur Internet, des élèves de Cm2 avaient tous retenu le nom « Google », même si d'autres moteurs avaient été présentés, au point que certains le voyaient également comme un « navigateur » ! Le moteur américain semble être parvenu à incarner Internet. D'ailleurs la supériorité du moteur Google est même approuvée par de nombreux professionnels de l'information² puisque près de 84 % plébiscitent Google.

Google devient ainsi le port à partir duquel beaucoup d'élèves avancent en pays inconnu. Dès lors, il est fort logique que les stratégies de recherche des élèves manquent de cohérence. Cette ignorance de l'objet technique Internet, de son histoire et de ses évolutions explique beaucoup les erreurs commises par les élèves. Ces derniers ne sont pas des Christophe Colomb et la serendipité (Ertzscheid 2003) n'est de fait guère fréquente. Il leur faut apprendre un nouveau langage car le moteur effectue la médiation entre l'utilisateur et le document, la traduction entre les intentions de recherche de l'utilisateur et le document potentiellement adéquat. L'entremise du moteur entre l'utilisateur et le document implique donc des usages performants et notamment l'emploi de mots-clés efficaces. Il y aurait donc un parler « moteur » pour ne pas dire un discours moteur combinant mots-clés et opérateurs booléens. Or ces pratiques ne sont pas naturelles et guère évidentes pour les jeunes usagers qui ont beaucoup de mal à définir ou à catégoriser. Les mots-clés impliquent une réflexion, une tentative pour résumer et définir une question en quelques mots. Il s'agit d'un langage différent du naturel. Or le langage naturel est présent, sous des formes certes parfois évoluées, dans d'autres applications sur Internet, notamment dans la messagerie ou les dialogues en direct.

2. La confusion des discours et les négligences.

L'Internet n'est pas uniforme, par conséquent les discours rencontrés diffèrent fortement suivant les actions effectuées. La communication médiatisée par Ordinateurs génère de nouveaux modèles d'interactions. Nous n'évoquerons pas ici les langages informatiques de l'internet même s'il nous apparaît important d'en montrer l'existence aux élèves. L'usage de l'affichage du code source peut s'avérer en effet riche en informations. Les discours et documents sur Internet connaissent de fortes variations suivant qu'il s'agit de « dialogues en directs », de blogs de type « journal intime », de sites d'informations, de messageries ou bien encore de flux de type Rss.

Les grilles élaborées par des documentalistes tentent de donner des pistes d'analyse de sites web aux élèves. Elles sont de plus en plus affinées et peuvent constituer un exercice intéressant. Bien souvent les élèves ne font pas attention à la validité d'une information car ils ne lisent pas tout (ce que les enseignants nomment parfois « l'effet zapping ») et surtout ils passent outre certaines indications. Ces mauvaises pratiques de lecture et d'identification font partie de ce que nous avons appelé les négligences.

Nous avons observé toutes les actions qui font que la liaison document-élève n'aboutit pas toujours au résultat escompté. Pour cela nous avons utilisé le concept de « négligences » qui définit tous ces phénomènes de non-lecture ou de mauvaise lecture. Cela nous a permis de définir une catégorisation des risques d'échec de la relation élève-document.

Le mot négligence vient du latin *negligentia* qui a un sens proche de l'acception actuelle. *Negligentia* vient de *negligere* qui se décompose en *neg-legere* qui signifie « ne pas lire » Nous avons

¹ Nous utilisons ici le terme pertinence dans son sens premier et non dans le sens que le moteur Google lui a attribué en le faisant devenir synonyme de popularité.

² Sondage effectué sur « le guide des égarés ». Résultats à la date du 14 mai 2006. [http://gde.jexiste.fr/joomla/Joomla_1.0.4-Stable-fr/component?option=com_poll/task/results/id,15/]

trouvé cette origine chez Régis Debray (Debray 1993) Les négligences sont donc par extension toutes ces actions de non-lecture, refus de lecture ou de « mauvaise » lecture. Ces actions négligentes produisent de nombreux effets néfastes que sont par exemple la mauvaise interprétation ou identification du document. Le faux ou l'erreur n'est donc pas ici intrinsèque au document mais extrinsèque. Le document qu'il soit numérique ou matériel voit son existence et sa distinction liées au fait qu'il faut que l'individu soit à même de le déchiffrer et de le comprendre. Or c'est bien souvent là que commencent les difficultés pour les élèves. L'emploi du mot « document » suscite déjà des difficultés. L'élève ne fait pas toujours nettement la différence entre un document et un documentaire. En clair, il ne perçoit pas souvent les limites du document qu'elles soient physiques ou sémantiques. Une difficulté d'appréhension du document d'autant plus complexe quand il s'agit d'effectuer des recherches sur Internet.

Toutes ces relations de négligences ne sont donc pas neutres et sans effet pour le document numérique. Nous avons dénombré six situations qui démontrent que la multiplication des sources d'information et les facilités d'accès qui l'accompagnent ne sont pas garants de la réussite documentaire :

- 1. Le document est ignoré :
- 2. Le document est plagié.
- 3. Le document n'est pas pertinent.
- 4. Le document est mal compris ou mal interprété.
- 5. Le document est incompréhensible.
- 6. Le document est inaccessible.

Nous sommes face à des processus communicationnels qui sont sources d'incompréhension. Alors que le principe de base d'Internet repose sur les possibilités offertes par l'hypermédia, nous remarquons un manque de lien entre l'utilisateur et le document. Le risque d'entropie n'est donc pas seulement lié à la multiplication des sources mais aussi à des barrières sémantiques voire techniques. Ces risques d'échec dans l'accès à la pertinence documentaire demeurent présents avec l'évolution constante des NTIC. Il en résulte des mutations documentaires qui compliquent l'identification des discours.

3. Information literacy et maîtrise de l'information.

Il faut se poser la question du devenir des médiateurs dans cette redocumentarisation. Ils vont de moins en moins constituer d'intermédiaires entre l'utilisateur et le document avec les difficultés évidentes qui en résultent. Le travail du RTP-Doc souligne bien que cette facilité d'accès apparente nécessite des savoirs et de nouveaux savoirs. La question mérite d'être posée : qui va se charger de la transmission de ces savoirs ? Cette tâche impliquerait une redéfinition des tâches professorales et un développement de la maîtrise de l'information au sein des systèmes éducatifs. Ou bien doit-on laisser l'utilisateur s'auto-former avec le risque d'une formation d'une « caste d'initiés » comme ils le sont qualifiés par Pédaque. L'« information literacy » apparaît alors comme une thérapie (Watzlawick 1979) pour résoudre le problème des négligences. Un travail important au niveau de l'évaluation de l'information numérique mérite d'être approfondi et organisé afin de donner aux usagers les capacités d'analyse et de critiques face aux flux rencontrés. Finalement il semble qu'il faille plaider pour un métadiscours, un discours de la méthode qui permettrait aux usagers d'acquérir leurs habiletés documentaires (« information literacy skills »). Les travaux et les réflexions sur ce sujet se multiplient à l'international. D'ailleurs l'IFLA (International Federation of Libraries Associations) vient de mettre en place une base de données qui regroupe les sites et travaux sur le sujet³. Cette volonté de prôner un discours méthodique et rigoureux s'observe dans la sémantique appliquée au sujet. Ainsi les anglosaxons n'hésitent pas à employer les termes de « grammar of the internet ».

Les stratégies des schémas heuristiques (mindmapping) constituent des pistes à exploiter en ce qui concerne l'éducation à la réflexion et à l'abstraction. Cela peut constituer de bons moyens pour que l'élève utilise les mots-clés avec efficacité. D'ailleurs certains moteurs comme Exalead offrent la possibilité de relancer la recherche à partir d'autres mots-clés affichés à partir d'une requête. Nous songeons aussi au métamoteur Kartoo qui par sa vision cartographique sort des habituelles démarches linéaires rejoignant quelque peu le mindmapping. L'élève prend ainsi l'habitude de travailler aussi sur les liaisons entre les concepts, les idées et les mots-clés. Il doit donc créer du lien avant même d'être connecté. Ces démarches « cognitives » vont dans le sens de l'augmentation de l'intellect via l'interaction Homme-Machine (Englebart 1963) Pour combattre les négligences et afin

³ International Information Literacy Ressources Directory. < http://www.uv.mx/usbi_ver/unesco/>

que l'élève puisse mieux appréhender les diverses formes de discours une voie semble devoir être développée : celle de la production de contenu par l'élève lui-même. La maîtrise des différents discours ne peut pas s'apprendre que par la lecture. Il faut donc inciter les élèves à devenir producteurs d'informations et à utiliser les folksonomies avec l'indexation par « tags » afin de mieux comprendre l'importance des mots-clés. C'est d'ailleurs tout le sens de l'Internet comme hypermedia et média « all-to-all ». Dès lors la réactivité face à l'information est préconisée et engendre échanges, commentaires, critiques et débats.

Conclusion :

Les possibilités offertes par les TAL, les systèmes de résumés automatiques ou bien encore de traduction de document ne sont pas suffisants pour garantir une pertinence documentaire. La relation usager-document aboutit à des échecs communicationnels du fait de négligences ou de discours non maîtrisés. La difficulté ne va faire que s'accroître avec l'hybridation des types de discours rencontrés sur une même page web. De ce fait l'accès à la pertinence documentaire ne peut se faire sans collaboration. Nous y voyons trois conditions pour que l'accès au document soit efficace :

- 1. La formation à l' « information literacy » doit se développer et être accessible à tous les usagers.
- 2. Les technologies doivent continuer à progresser afin que les robots puissent indexer de manière plus performante les différentes données. Les moteurs doivent évoluer notamment de manière à distinguer pertinence et popularité.
- 3. L'hybridation des deux premières conditions trouve son prolongement dans le développement de la troisième condition : mieux intégrer l'interaction avec l'utilisateur.

Dès lors le cercle vertueux peut se mettre en place mais ce succès ne peut avoir lieu sans rapprochement des techniques et des usages. Des métadonnées efficaces pourront être émises et le projet de web sémantique pourra peut-être dès lors voir le jour. Il faut pour cela que tout le monde travaille dans le même sens. Par conséquent le rôle de médiation des professionnels de l'information doit s'accroître au niveau de la formation ce qui implique sans doute des mutations professionnelles.

Bibliographie :

- DEBRAY Régis (1993). *Vie et mort de l'image. Une histoire du regard en occident*. Paris : Gallimard, 1993.
- DUMAS Philippe (2005) Google au quotidien : le googling ou les habitudes de recherche de l'internaute ordinaire. Communication, *Le Monde selon Google*, Bucarest, Juin 2005. http://archivesic.ccsd.cnrs.fr/sic_00001577.html
- ENGLBART Douglas. (1963) A Conceptual Framework for the Augmentation of Man's Intellect in *The Augmentation of Man's Intellect by Machine*, vol. 1, Howerton and Weeks, Eds. Washington, DC: Spartan Books. p 1-27.
- ERTZSCHEID Olivier., GALLEZOT, Gabriel. (2003) "Chercher faux et trouver juste : sérendipité et recherche d'information. 1ère conférence internationale francophone en Sciences de l'Information et de la Communication 10ème colloque bilatéral Franco-Roumain. Bucarest. Juillet 2003. <http://archivesic.ccsd.cnrs.fr/documents/archives0/00/00/06/89/sic_00000689_02/sic_00000689.html>
- RTP CNRS 33 <RTP-DOC> (2005-2006) Documents et contenu : création, indexation, navigation. *Plateforme d'échange du Réseau thématique pluridisciplinaire sur le document* numérique. <<http://rtpdoc.enssib.fr>>
- PROULX Serge (2001) « Usages des technologies d'information et de communication : vers une reconsidération du champ d'étude », conférence à *Inforcom 2001*, Congrès de la Société des sciences de l'information et de la communication, Paris. <http://grm.uqam.ca/textes/proulx_SFSIC2001.pdf>
- SERRES Alexandre (2005) Évaluation de l'information sur Internet : Le défi de la formation, *BBF*, n° 6, p. 38-44 <<http://bbf.enssib.fr>>
- SIX DEGRES, LE SPHINX DEVELOPPEMENT. (2006) *Les usages de l'Internet dans l'enseignement supérieur : de la documentation au plagiat* < http://www.compilatio.net/files/sixdegres-sphinx_enquete-plagiat_3fev06.pdf>
- WATZLAWICK Paul. (1979) Une logique de la communication. Le seuil. Points essais.