
Apprendre la recherche d'informations sur Internet à l'école : à la découverte du poids des mots

Congrès international AREF 2007 (Actualité de la Recherche en Education et en Formation)

Caroline Ladage

UMR ADEF

Université de Provence

1 rue de Verdun

F-13410 LAMBESC (AIX)

caroline.ladage@univ-provence.fr

RÉSUMÉ. Alors que la recherche d'informations sur Internet constitue une des activités principales effectuées sur Internet, les pratiques de recherche d'informations semblent souffrir d'un manque de diffusion et de didactisation. Cet article s'inscrit dans une recherche qui a pour objectif d'étudier la problématique de sa diffusion, ainsi que les savoirs utiles à son enseignement. Partant de l'observation, à l'école primaire, de séances d'enseignement consacrées à la formulation de requêtes sur des moteurs de recherche, nous tâchons de mettre en évidence l'importance de l'étude des praxéologies de recherche d'informations à l'école et de leur analyse didactique. Nous explorons, dans le cadre de la théorie anthropologique du didactique, le concept de dialectique des médias et des milieux comme référent théorique pour penser et repenser l'enseignement de la recherche d'informations sur Internet dans une perspective globale.

MOTS-CLÉS : recherche d'informations, moteurs de recherche, didactique, médias et milieux.

Introduction

Depuis la mise en place du B2i (Brevet Informatique et Internet, niveaux primaire et collège) et du C2i (Certificat Informatique et Internet, niveau Lycée), la recherche d'informations sur Internet doit faire partie intégrante des apprentissages à tous les niveaux scolaires. Constatant que cette pratique est loin d'aller de soi pour bon nombre d'utilisateurs du web, notre projet est d'explorer les connaissances nécessaires pour un bon usage des outils de recherche sur Internet en vue d'approfondir les approches didactiques actuelles, voire d'en modifier la perspective, afin d'intégrer son enseignement dans une approche didactique globale. Cette étude s'inscrit dans une recherche plus large incluant des observations au collège, au lycée et à l'université.

À partir d'observations réalisées en classe de CM2 (Cours Moyen 2^{ème} année) de l'école primaire, lors de séances organisées pour l'apprentissage des compétences visées au B2i¹, nous explorons dans cet article plus particulièrement les conditions et contraintes de l'enseignement de la recherche d'informations sur Internet à l'école primaire. Constatant les difficultés des élèves de tous les niveaux à formuler des questions de recherche et à les exprimer en requêtes sur les outils de recherche utilisés, nous explorons les processus spontanés de définition des requêtes en situation de classe et évaluons le poids des types de requêtes pour une recherche d'information réussie.

1. Activités de recherche d'informations dans une classe de CM2

Le B2i école² vise l'acquisition d'une série de connaissances, capacités et attitudes comme fondement de la mise en œuvre de compétences permettant aux élèves, dans le domaine de la recherche d'informations qui nous intéresse ici, de « s'informer et de se documenter » à l'aide d'outils de recherche sur Internet. Les différentes séances que nous avons observées avaient ainsi pour objectif d'approfondir les connaissances des élèves sur la recherche d'informations sur Internet (RII) et étaient consacrées à l'utilisation des moteurs de recherche. La

¹ Bulletin officiel n° 29 du 20 juillet 2006.

² La recherche d'informations sur Internet fait l'objet, dans le référentiel du B2i école, du domaine 4, « s'informer et se documenter », dont les contenus sont définis de la manière suivante : *Connaissances principales* : Les outils de recherche utilisent des critères de classement et de sélection de l'information. *Objectif* : Lire un document numérique. Chercher des informations par voie électronique. Découvrir les richesses et les limites des ressources de l'Internet. *Capacités* : l'élève doit être capable de consulter un document à l'écran ; d'identifier et trier des informations dans un document ; d'utiliser les fonctions de base d'un navigateur ; d'effectuer une recherche simple.

première séance avait pour objectif de faire découvrir aux enfants les effets de la polysémie d'un mot sur les résultats des requêtes d'un moteur de recherche. La seconde demandait aux élèves un travail de formulation autonome de requêtes pour trouver des réponses à une question donnée : « *pourquoi Pluton n'est plus considéré comme une planète ?* » La troisième proposait le même travail, en l'abordant cette fois en groupe. Nous nous proposons d'approfondir plus particulièrement les résultats de la seconde séance en analysant les requêtes formulées par les élèves.

2. Analyse des historiques des pages visitées et des mots clés utilisés

L'analyse des historiques des requêtes a pour objectif d'explorer la formulation des requêtes adressées aux moteurs de recherche par les élèves.

2.1 Méthode de recueil des données

Afin de recueillir les requêtes³ formulées par les élèves, nous avons procédé à l'enregistrement des adresses web (URL⁴) des moteurs de recherche dans les historiques des navigateurs (exclusivement Internet Explorer) utilisés pendant la séance durant laquelle les élèves avaient fourni un travail autonome de recherche sur Internet. Nous avons ainsi recueilli les historiques de 9 sessions, c'est-à-dire le travail de 18 élèves, étant donné qu'ils travaillaient en binômes. À partir de ces historiques nous avons retenu 95 requêtes pour notre analyse⁵.

2.2 Identification de types de requêtes

L'accès à l'information en ligne et la question de la définition des requêtes est une problématique étudiée dans différents champs disciplinaires, tantôt pour étudier les comportements cognitifs des usagers des systèmes d'information (Tricot, 2003), tantôt pour améliorer la recherche d'information grâce à certains langages de programmation (dont le XML, ou Extensible Mark-up Language (Englmeier, Hubert, et Mothe (2006)). Rappelons aussi que la problématique de l'étiquetage de l'information en lien avec la recherche d'informations sur Internet est au cœur du

³ Par « requête » nous entendons les mots clés et expressions écrites dans les fenêtres de recherche des outils de recherche.

⁴ L'adresse web ou URL (Uniform Resource Locator) d'une page est généralement enregistrée dans l'historique du navigateur. La particularité d'une page visitée sur un moteur de recherche est que son URL contient les mots clés de la requête formulée par l'utilisateur.

⁵ Nous ne retenons pas l'occurrence des termes de recherche sur un même ordinateur, pour un même outil de recherche. Nous retenons leurs occurrences sur un même ordinateur mais sur des outils de recherche différents, ainsi que sur des ordinateurs différents.

domaine de recherche du traitement automatique du langage naturel (Véronis, 2000), ainsi que du Web sémantique (Shadbolt, Berners-Lee, Hall (2006)).

Connaissant les différences d'impact de chaque type de requête sur la qualité des résultats qu'un moteur de recherche renvoie, nous essayons d'identifier les différents types de requêtes formulés par les élèves, afin de mieux cerner les difficultés qu'entraînent l'utilisation de chacune d'entre elles et d'offrir des éclairages, voire de proposer d'autres types de formulation de requêtes, afin de rendre leurs RII plus performantes. Les critères que nous avons retenus pour le repérage de catégories sont les suivants (suivi du pourcentage de requêtes à l'intérieur des différentes catégories identifiées dans notre corpus de requêtes retenues) :

- formulation d'une question (voir le tableau 1 ci-dessous),
soit 27 % des requêtes ;
- expression à plusieurs termes avec une structure de phrase,
soit 12 % des requêtes ;
- expression recopiée à partir de la fiche de questions,
soit 5 % des requêtes ;
- expression simplifiée à trois ou quatre termes (Pluton plus planète),
soit 4 % des requêtes ;
- requête d'un seul terme (planète (24 %) ; Pluton (17 %)),
soit 41 % des requêtes ;
- requête de deux termes (planète Pluton),
soit 11 % des requêtes.

Dans la situation de classe que nous étudions ici, nous n'avons bien sûr pas rencontré tous les types de requêtes qui peuvent être utilisés pour interroger un moteur de recherche, et, eu égard aux évolutions des technologies et aussi des différents types de contenus en ligne (texte, audio, vidéo, image, base de données), cet inventaire de types de requêtes n'est qu'une ébauche appelée à évoluer. Ces requêtes couvrent la recherche simple tel quelle est visée par le B2i, et ne comprennent pas les requêtes d'un niveau plus avancé que nous étudions en travaillant avec des élèves du collège ou du lycée, ainsi qu'avec des étudiants ou des professionnels. Enfin il va sans dire que les différents types de requêtes ne relèvent pas de l'usage exclusif de certains groupes d'utilisateurs.

2.3 Les requêtes : de la question complète au mot clé unique

Les questions représentent 27 % du nombre total des expressions recueillies. Seul un binôme ne formule pas de question. Si on cumule ce résultat avec les 12 % des « autres expressions » et les 5 % des « expressions recopiées », totalisant ainsi 44 % des expressions, on voit qu'une part importante des élèves recourt aux

requêtes longues, que ça soit une expression complète, une phrase ou une question. Les requêtes simples, « planète », « Pluton », « planète Pluton », représentent 52 % du nombre total des expressions. En observant ainsi la longueur des requêtes nous constatons une situation aux deux extrêmes : soit des requêtes longues (jusqu'à dix termes), soit des requêtes courtes (un ou deux termes), mais rarement des requêtes que nous avons appelées « simplifiées », de trois ou quatre termes, qui de surcroît, pendant les séances que nous avons observées, n'étaient pas utilisées spontanément, mais uniquement à la suite de l'intervention de l'enseignante.

Sans entrer ici dans le détail du fonctionnement des moteurs de recherche, regardons d'un peu plus près ces différents types d'expressions pour mieux comprendre l'impact de chaque type sur les résultats que renvoient ces outils.

2.3.1 Les requêtes courtes

Les requêtes courtes, d'un seul mot, représentaient 41 %. Elles étaient rarement fructueuses par rapport aux réponses cherchées par les élèves. La requête « Pluton » renvoyait ainsi à la fois à des sites qui le considéraient (encore) comme une planète, et à des sites qui annoncent déjà son rejet du rang des planètes, la requête « planète » ne mentionnant même pas « Pluton » sur ses deux premières pages de résultats. Il est en effet difficile à partir d'un seul mot clé, d'obtenir une réponse pertinente. En recherche documentaire on distingue généralement les mots clés « génériques » (pas de spécialisation de domaine) et les mots clés « spécialisés » (le domaine est fixé). La connaissance de la polysémie des mots, du degré de spécificité ou de généralité d'un mot facilite donc la formulation d'une requête. Pour sensibiliser les élèves à ces phénomènes, la première séance avait d'ailleurs portée sur la recherche de toutes les catégories de sites et d'informations qu'on pouvait trouver sur le mot clé « air ».

2.3.2 Les requêtes longues

Les requêtes longues comprennent dans notre corpus aussi bien les questions, les expressions recopiées et les phrases. Si on regarde de plus près les questions posées dans le tableau 1 ci-dessous, on constate qu'il s'agit de variations sur les questions posées par l'enseignante⁶, comme si les élèves faisaient un transfert de la question à une tierce personne. Certains font des efforts de simplification, d'autres formulent des questions complètes et vont jusqu'à mettre un point d'interrogation. La plupart sont marquées par la présence de mots interrogatifs (pourquoi, quand, qui).

Requêtes d'élèves sous forme de questions
--

⁶ Le travail de recherche à faire était de répondre aux questions suivantes : « Depuis quelque temps, Pluton n'est plus considérée comme une planète. Quelles en sont les raisons ? Quand cela a-t-il été décidé ? Qui l'a décidé ? Quelle est la carte d'identité de cette « planète » ? »

pluton planete ou pas, / quand pluton na plus etait une planete / quelle est la carte d'identité de pluton / pourquoi pluton n est plus une planete / pourquoi pluton n est plus une planete wikipedia / quand pluton n est plus une planete / pourquoi pluton est plus considérée comme une planète / pourquoi pluton plus planète ? / pluton est t'elle une planete / pourquoi pluton n'est plus une planète ? / pluton est t'elle une planete / pluton n'est plus une planete pourquoi ?? / quand cela à t-il était que Pluton devait quitter la carte / quand cela a t-il été décidé que Pluton n' est plus considérée comme une planète / De puis combien de temps Pluton ne fait plus parti du système solaire / pourquoi Pluton n'est plus considerée comme une planète / Quand on a décidé que la planete pluton ne faisait plus parti du système solaire / Qui a décidé que Pluton ne faisait plus parti du système / quand la planète Pluton ? / quand la planète Pluton a etait non considéré comme planète ?
--

Tableau 1. *Les requêtes sous forme de questions*⁷

Dans notre observation les requêtes longues étaient, elles aussi, rarement fructueuses et ont poussé les élèves à la reformuler, passant alors souvent à la formulation d'un seul mot clé. Une requête longue sur un moteur de recherche ne se justifie réellement que dans certains cas, comme la recherche de phrases exactes ou de titres d'œuvres. Généralement on utilise l'opérateur booléen⁸ des guillemets pour signifier qu'on est à la recherche de la phrase exacte. En lançant une requête longue on trouve toujours quelque chose en rapport avec le sujet recherché, mais on y perd aussi souvent du temps.

2.3.3 *Les requêtes simplifiées*

Par « requête simplifiée » nous entendons les requêtes, généralement de plus de deux termes, qui n'ont pas obligatoirement une structure de phrase mais qui comprennent les termes essentiels, représentant au mieux une question de recherche. Dans la question qui était posée lors de la séance observée, les élèves pouvaient par exemple explorer la requête « Pluton plus planète ».

2.4 *La difficile quête du mot juste*

On observe dans l'ensemble une formulation de requêtes plutôt chaotique et coûteuse. Il ne va en effet pas de soi pour des élèves de CM2 de trouver les termes adéquats par rapport à une question de recherche un peu complexe. Par rapport au

⁷ Nous avons laissé les fautes d'orthographe, mais enlevé les doublons.

⁸ Un opérateur booléen est une commande logique (de la logique de Boole, d'après le mathématicien anglais du 19^e siècle George Boole). Les opérateurs booléens sont très utilisés pour écrire des équations de recherche lors de l'interrogation des banques bibliographiques ou du Web (*Dictionnaire de l'information* (2006). Armand Colin).

fonctionnement des moteurs de recherche, si chaque élève a compris les principales fonctionnalités (fenêtre de recherche, bouton pour lancer la requête, pages de résultats avec liens à cliquer), ils ne comprennent pas *comment* l'information est traitée, ni *qui* la traite. On voit également à l'œuvre un phénomène de personnalisation de l'outil, illustré par des remarques d'élèves du type « *on peut lui dire ça comme ça ?* », « *Eh ! Tu peux mettre "pluton plus planète", il comprend* ».

Les oscillations que nous avons observées entre requêtes longues et requêtes courtes témoignent du manque de compréhension du traitement automatique du langage propre aux moteurs de recherche.

Cet aperçu des résultats invite à réfléchir sur la manière dont on peut amener de jeunes élèves à apprendre à cerner des domaines de connaissances et à choisir les mots clés et expressions adaptés. Par la comparaison des requêtes de tout un chacun au sein du groupe, les élèves pourraient ainsi être encouragés à revoir leurs requêtes et soit les simplifier, soit les enrichir. La lecture des pages de résultats des moteurs de recherche devrait, si la requête ne renvoie pas des informations pertinentes, toujours inviter les utilisateurs à reformuler leurs requêtes, ce qui est encore loin d'aller de soi pour nombre d'entre eux.

Notre exploration des types de requêtes a ainsi pour objectif de mettre en évidence des pratiques d'élèves. Elles peuvent servir de clé de compréhension et d'explication du processus de formulation des requêtes d'une part, et de leurs résultats dans les outils de recherche d'autre part. Pour aller plus loin, soulignons que l'incidence de l'utilisation de chaque type de requête ne dépend pas seulement du contexte dans lequel les mots clés utilisés sont susceptibles de s'inscrire. Elle dépend aussi de l'usage qui est fait de l'information recueillie. Enfin, la question des apprentissages de la formulation et de la compréhension des requêtes est d'autant plus importante lorsqu'on sait les effets de l'optimisation de contenus sur Internet en vue de leur référencement. L'optimisation de contenus pour les moteurs de recherche a pour effet, pour un part importante de documents en ligne, de fausser les algorithmes de classement des moteurs de recherche, et de présenter pour nombre de requêtes, des contenus commerciaux, politiques ou moralement discutables. Il s'agit là de questions qui vont bien au-delà des enseignements actuels de la RII et qui nécessiteraient une approche globale que nous proposons de clarifier à la lumière de la théorie anthropologique du didactique (Chevallard, 2006).

3. Une culture à construire

Les résultats que nous avons présentés dans cet article ne sont qu'une facette d'un phénomène, qui, avec les évolutions rapides des technologies de l'information, devient un ensemble toujours plus complexe de praxéologies (au sens de Chevallard (2003)), encore peu et/ou mal diffusées et très peu didactisées. En RII nous constatons qu'il y a des pratiques pour lesquelles il existe presque exclusivement des praxéologies personnelles, qui ne sont pas partagées. Il n'est pas facile d'observer

ces praxéologies, chacun les gardant un peu pour soi. Cette résistance à expliquer les techniques utilisées débouche le plus souvent sur une dénégation du didactique. Dans l'absence de partage, on ramène alors les pratiques de recherche à une seule façon de faire, que les manuels d'utilisation d'Internet ne manquent pas de répéter, sans toutefois qu'une analyse praxéologique soit faite. Le type de tâches et la technique pour l'accomplir, au lieu d'être distincts, se trouvent ici soudés. On retrouve cette confusion dans l'injonction maintenant bien connue des élèves : « Regardez sur Internet ! » Si la tâche est facile, ça va ; mais si elle est difficile cela pose problème, par manque de techniques : l'outillage manque en effet souvent pour affronter une tâche d'un certain niveau de difficulté.

Dans le domaine de la RII il n'y a ainsi pas de standardisation des techniques : on est en présence d'un univers praxéologique flou. Mais ce manque de diffusion et de standardisation ne viendrait-il pas aussi du fait qu'une technique qui se serait montrée efficace un jour peut souvent s'avérer inefficace dès lors qu'on essaie de la mettre en œuvre dans d'autres types de recherches ? Pourquoi alors l'enseigner, pourquoi même la retenir ?

Pour dépasser ces impasses un autre regard devrait être porté sur la question de la pertinence et de l'efficacité des techniques de recherche. Une telle problématique participe du constat énoncé par Chevallard : « *l'un des grands problèmes éducatifs et citoyens est celui de la généralisation de la capacité (de l'élève, du professeur, du formateur, du chercheur, du citoyen, etc.) à situer sa pensée et son action dans une dialectique des médias et des milieux adéquate à l'évaluation de ses assertions et de ses décisions* » (Chevallard, 2007).

Une technique de recherche devrait être diffusée, enseignée, non comme une technique immuable et standardisée, applicable en toute circonstance, mais comme une technique qu'il faut confronter aux milieux dans lesquels elle s'inscrit et dans lesquels elle produit des effets. Ce qui s'entend dans le cas que nous étudions ici par : un certain type de requête, dans un moteur de recherche donné (au fonctionnement, au passage, opaque), explorant une fraction de contenus indexés à un instant *t*, appelant des domaines de connaissances multiples non catégorisés, à l'envergure inconnue a priori. Internet est caractérisée par une forte évolutivité et une forte densité de médias et de milieux : aussi il faut, quand on s'y aventure, en permanence vérifier, non pas seulement la validité des informations qu'on y trouve, mais aussi la pertinence et la capacité de nos techniques pour les trouver. Il ne s'agit en RII en effet pas simplement d'apprendre à manipuler un outil : tout un ensemble de praxéologies entrant en jeu pour outiller l'utilisateur à la recherche d'informations, en établir la liste a priori reviendrait à se fermer, et à croire qu'il n'y aurait là rien à « apprendre ». Ainsi au lieu de s'en tenir à la diffusion d'astuces pour dénicher de bonnes informations, la RII peut, à notre sens, bel et bien faire l'objet d'enseignements à l'école et à l'université et dépasser le cadre des pratiques individuelles pour être étudiée ensemble. Pour apprendre cette recherche d'informations-là il faut faire entrer ces pratiques dans les classes et les intégrer dans l'ensemble des apprentissages scolaires, dans une perspective plus large,

intradisciplinaire et codisciplinaire. La RII, en mathématiques, en français, en histoire, en philosophie, en économie, etc., peut se décliner en autant de domaines de connaissances qu'il y a de questions à se poser et à se reposer dans notre société actuelle. Cela en même temps contribuerait certainement, non sans heurts, à faire sortir l'école de son enfermement. Il va de soi qu'il ne s'agit pas là de reproduire un cloisonnement disciplinaire, un « *retrait sur soi des disciplines* » (Chevallard, 2003), mais au contraire de permettre, de se donner la liberté de regarder comment chaque discipline propose de répondre aux questions posées, en étant conscient qu'informations et techniques doivent être questionnées dans une dialectique entre médias et milieux.

4. Questionner la requête et l'information

Souvent les exercices en RII se limitent à des requêtes précises et/ou imposées, données a priori, ce qui ne laisse pas aux élèves la liberté d'en découvrir d'autres, par eux-mêmes, en cours de recherche, d'explorer les pages pour trouver des termes permettant d'affiner leurs requêtes. En RII il est important de développer une culture du questionnement de l'information, des médias et des milieux dans lesquels on la cherche et où on la trouve, d'apprendre à cerner un domaine de connaissances, de devenir en quelque sorte un explorateur du Web. C'est par un effort d'exploration à partir d'un moteur de recherche et sur les pages web que ce moteur propose qu'on découvre les différents milieux dans lesquels les termes qu'on expérimente ont du sens. Pour certains termes on trouve, comme les élèves de CM2, un nombre important de significations. C'est en affinant sa requête qu'on enrichit les résultats de sa recherche. C'est en mettant en place une réflexion sur la dialectique des médias et de milieux que nous serons capables de déterminer la validité des informations proposées. Quels médias pour quel type de recherche ? Quels milieux pour répondre à telle question de recherche ? Quelles requêtes pour y accéder ? Ainsi une requête sur le statut de planète de Pluton ne saurait se contenter de l'information diffusée par un élève qui raconte sa journée d'école sur son blog. Si toutefois cette requête avait pour objectif de cerner la représentation des élèves face à la perte du statut de planète de Pluton, la même page de blog aurait été parfaitement adéquate.

Notre observation en classe avait ainsi pour objectif d'explorer les conditions et contraintes d'un enseignement en RII qui ne se limite pas exclusivement à une revue des fonctionnalités des outils de recherche, mais qui ait pour projet de faire découvrir aux élèves, dès l'école primaire, le poids des mots comme autant de clés pour accéder aux informations en ligne.

5. Bibliographie

Brevet Informatique et Internet (2006). Bulletin officiel N° 29 du 20 juillet 2006. [En ligne]

10 Actualité de la Recherche en Education et en Formation, Strasbourg 2007

<http://www.education.gouv.fr/bo/2006/29/MENE0601490A.htm>. (Consulté le 26.03.2007).

Chevallard, Y. (2007). *Journal du séminaire didactique des savoirs mathématiques pour formateurs*. [En ligne] <http://www.aix-mrs.iufm.fr/formations/filieres/mat/dfd/2006-2007/excursus07.html> (consulté le 26.03.2007).

Chevallard, Y. (2003). *Didactique et formation des enseignants*. Communication aux Journées d'études INRP-GÉDIAPS Vingt ans de recherche en didactique de l'Éducation Physique et Sportive à l'INRP (1983-2003) (Paris, 20 mars 2003). Paru dans B. David (éd.), *Impulsions 4*, INRP, Lyon, 2005, p. 215-231. [En ligne] http://yves.chevallard.free.fr/spip/spip/article.php3?id_article=63. (Consulté le 26.03.2007).

Englmeier, K., Hubert, G., Mothe, J. (2006). Distinguer les requêtes pour améliorer la recherche d'information XM. Actes du colloque CORIA 2006. pp. 41-52. [En ligne] <http://www.irit.fr/ARIA/2006/41.pdf>. (Consulté le 23.03.2007).

Shadbolt, N., Berners-Lee, T., Hall, W. (2006). The Semantic Web Revisited, *IEEE Intelligent Systems* 21(3) pp. 96-101, May/June 2006. [En ligne] http://eprints.ecs.soton.ac.uk/12614/01/Semantic_Web_Revisted.pdf. (Consulté le 23.03.2007).

Tricot, A. (2003). *Apprentissage et recherche d'information avec des documents électroniques*. Mémoire en vue de l'habilitation à diriger des recherches. ERT « hypermédias et apprentissages », Toulouse, 136p.

Véronis, J. (2000). *Sense tagging: Don't look for the meaning but for the use*, Computational Lexicography and Multimedia Dictionaries (COMLEX'2000) (pp. 1-9). Kato Achia (Greece). [En ligne] <http://www.up.univ-mrs.fr/~veronis/pdf/2000complex.pdf>. (Consulté le 26.03.2007).