

Pratiques documentaires en sciences avec Internet, outils de recherche et compétences des élèves

François-Marie Blondel

*Institut National de Recherche Pédagogique
Département Technologies Nouvelles
91, rue Gabriel Péri
92120 Montrouge
Francois-Marie.Blondel@inrp.fr*

Résumé

L'objet de l'étude présentée dans cet article est d'analyser la manière dont les élèves de lycée utilisent le web pour des activités de recherche documentaire en sciences. Trois classes ont été observées et les traces de leur navigation enregistrées. Les élèves éprouvent quelques difficultés à utiliser les annuaires comme Yahoo et en tirent peu d'information en général. Ils préfèrent les moteurs de recherche comme AltaVista pour le grand nombre de résultats qu'ils renvoient, même s'ils n'en consultent pas toutes les adresses qui paraissent pertinentes. La formulation des requêtes et en particulier l'emploi de synonymes et leur combinaison pour construire une requête efficace reste une des difficultés majeures. Sur la base de ces résultats, nous étudions un prototype de navigateur qui facilite la planification de l'activité de recherche tout en fournissant des conseils sur les techniques à mettre en œuvre.

Abstract

The purpose of this study was to analyze how high school students use the World Wide Web to find information in the field of science. Three groups of students were observed during their searching activities and the tracks of their navigation were recorded. The students show more difficulties in exploring directories like Yahoo and generally get few results from them. They usually prefer search engines like AltaVista because of the large amount of pages found, even if they are not capable to visit the most pertinent part of them. One of the major problems encountered by students relies in generating their search statements, finding synonyms and combining them in an efficient statement. Based on the results of this study, we are developing a dedicated browser, which includes guidance in search techniques and facilitates planning and control of the overall search activity.

Mots-clés

Recherche documentaire, science, internet, annuaires, moteurs de recherche

Keywords

Information retrieval, science, web tools, directories, search engines

1. Introduction

Internet offre en matière documentaire des quantités de ressources nouvelles difficilement comparables avec les ressources disponibles par les moyens habituels de documentation des établissements d'enseignement. En sciences, de nombreux sites publics et privés donnent accès à des données d'observation et de mesure mais aussi à des analyses et à des publications scientifiques sous des formes variées.

Si ces possibilités étendues en matière d'activité documentaire scientifique sont maintenant mieux identifiées, l'intérêt commence seulement à en être perçu par les enseignants et les usages éducatifs que l'on peut en attendre restent encore à définir. De surcroît, la manière dont les élèves utilisent et maîtrisent ces nouveaux outils est encore mal connue.

Dans le cadre d'une recherche qui s'intéresse aux dispositifs qui facilitent des activités d'apprentissage autonome des étudiants, nous avons entrepris une étude des outils disponibles pour des activités documentaires et une analyse de la manière dont les élèves s'approprient et maîtrisent ces outils pour conduire une recherche d'information sur un sujet scientifique.

Nous présentons ici les résultats de plusieurs observations effectuées sur des groupes d'élèves de lycée pratiquant ce type d'activités. Nous discutons ensuite plusieurs propositions destinées à faciliter ces usages documentaires.

2. Méthodologie d'observation et pratiques observées

Au cours de l'année scolaire 1998/99, nous avons effectué plusieurs observations de groupes d'élèves dans des activités de recherche d'information et de documentation utilisant Internet : une classe de Première Scientifique effectuant des recherches sur le nucléaire et la radioactivité (groupe HPS), une classe de BTS en Economie Sociale et Familiale sur les sources d'énergie à la maison (groupe MBTS), une classe de Première Scientifique sur les filières conduisant à la production d'énergie électrique (groupe FPS).

Avant la recherche sur Internet, les élèves ont bénéficié d'une courte initiation (de 10 à 30 minutes) portant sur les principaux annuaires et moteurs, leurs différences principales et leurs syntaxes de base.

L'observation était effectuée par une personne par classe. Le recueil des traces de navigation a été effectué au moyen d'un logiciel de traçage que nous avons développé. Ce logiciel conserve la date et l'adresse de chaque action effectuée sur le navigateur, mémorisant ainsi les appels aux moteurs et aux annuaires, la formulation des requêtes, et les adresses de toutes les pages visitées.

3. Premiers constats : des outils mal employés

Une analyse qualitative des observations et des traces de navigation conduit à une situation assez contrastée dans laquelle une grande aisance de manipulation côtoie une difficulté à employer efficacement les principaux outils disponibles : moteurs et annuaires. Néanmoins les résultats de ces recherches restent rarement infructueux, les élèves trouvant au moins quelques documents répondant grosso modo à la question posée.

Une aisance de manipulation

Les élèves n'éprouvent aucune difficulté à utiliser un navigateur et à manipuler les interfaces des principaux outils de recherche d'information. L'emploi simultané d'un navigateur et d'un traitement de texte, qui conduit à basculer rapidement d'une application à une autre, ne leur a pas posé de difficultés majeures.

Tout au plus, peut-on observer une réticence à gérer plusieurs fenêtres de recherche simultanées (cette facilité est rarement utilisée) ainsi que quelques lenteurs dues aux allers et retours permanents entre la page de résultats d'un moteur et la consultation des sites mentionnés dans cette page.

Une confusion entre annuaires et moteurs

Bien que la distinction entre ces deux types d'indexation de la toile leur ait été présentée, les élèves ne semblent pas bien en percevoir les différences. A de rares exceptions près, ils n'en tiennent pas compte pour rechercher des informations. Nous avons observé à plusieurs reprises qu'ils formulent exactement les mêmes requêtes en interrogeant un annuaire et un moteur. Ainsi, un élève du groupe FPS interroge-t-il successivement Yahoo puis Altavista avec les mots "cellules photovoltaïques".

De plus, ils ne semblent pas percevoir de différences dans les résultats fournis et en déduisent immédiatement que les moteurs sont plus efficaces, ces derniers produisant plus de résultats (en nombre de pages trouvées) que les annuaires (en nombre de sites trouvés). Ainsi l'élève cité précédemment obtient-il 3 réponses (qui d'ailleurs pointent sur le même site) avec Yahoo et plus de 36000 pages annoncées avec Altavista.

Il faut noter ici que cette distinction est rendue plus délicate à mettre en évidence par l'existence de sites portails qui abritent à la fois un annuaire et un moteur.

Des annuaires jugés peu utiles

La consultation des annuaires en partant des rubriques de classement hiérarchique, démarche peu fréquente, ne débouche que rarement sur la consultation de sites. Tout se passe comme si les élèves ne trouvaient pas leur chemin dans cette hiérarchie.

Par ailleurs, l'interrogation des annuaires par mots clés ne produit pas toujours de nombreux résultats (cf. supra) sans que pour autant les élèves ne remettent en cause leur manière d'interroger. Un élève du groupe HPS interroge Yahoo France avec 3 requêtes différentes sur les catastrophes nucléaires sans obtenir de résultats ; à la question posée par l'observateur "Tes requêtes ne produisent pas grand' chose", il répond "Si on pouvait trouver avec ces mots-là, ce serait parfait".

En règle générale, quand ils en ont le choix, les élèves préfèrent appeler un moteur plutôt qu'un annuaire. Un élève du groupe HPS interrogé sur ses préférences juge d'ailleurs Yahoo "trop compliqué" et Altavista "plus clair". Sur l'ensemble des groupes observés, nous avons constaté que le temps consacré à l'usage des moteurs l'emporte largement sur celui consacré à l'usage des annuaires.

Une difficile formulation des requêtes

La formulation des requêtes est un point important de la démarche de recherche d'information qui semble poser de réelles difficultés aux élèves.

Tout d'abord, bon nombre d'entre eux se contentent de recopier mot pour mot la question posée. Un élève du groupe MBTS qui devait rechercher les différents prix de l'énergie électrique utilisée à la maison dans toute l'Europe a tout simplement introduit sa requête sous la forme "énergie électrique à la maison et en Europe". Cette formulation dénote un manque de familiarité, assez généralement observé, avec les techniques d'indexation et d'interrogation de bases de données.

L'emploi de synonymes est aussi relativement limité, l'évocation de mots représentatifs de la question posée étant probablement une difficulté pour de nombreux élèves. Par exemple, un élève du groupe MBTS qui recherche sur les différents combustibles utilisés à la maison, a employé les formulations suivantes : "chauffage au fioul", "chauffage fioul", "combustible au fioul", "chauffage au fioul domestique", "fioul domestique".

Quand ils ne recopient pas la question, les élèves ont une forte tendance à employer le minimum de mots dans leur requête. Sur un ensemble de 52 requêtes formulées par 6 élèves représentatifs du groupe FPS, la moyenne se situe à 2 mots par requête, sans qu'il y ait de différence entre celles qui sont adressées aux annuaires et celles qui sont adressées aux moteurs. Sur ce total, 20% seulement comportent 3 ou 4 mots.

Si l'utilisation d'opérateurs booléens pour exprimer des requêtes plus fines est très rare (un seul élève sur tous les groupes observés), l'emploi d'opérateurs qui précisent l'interrogation est plus fréquent. Les élèves du groupe FPS ont employé l'opérateur + (pour imposer la présence d'un mot) dans 30% de

leurs requêtes et l'opérateur "" (expression entière) dans 20 % d'entre elles. Ces deux opérateurs présentent en effet l'avantage d'être faciles à interpréter.

Un long temps de recherche

Plusieurs observations nous ont montré que la part de l'activité consacrée à l'appel des outils, à la formulation de la requête et à la consultation des réponses, est importante, comparée à la part consacrée à la consultation des sites et des pages. Même si cette dernière reste majoritaire, les élèves consacrent au moins le tiers et le plus souvent la moitié de leur temps à formuler leurs requêtes et à consulter les réponses.

En mettant à part la lenteur de réponse de certains serveurs, cette durée s'explique aussi par la difficulté à lire et à interpréter les résultats affichés par les moteurs. Ceux-ci sont parfois accompagnés de descriptions sibyllines et l'interprétation des informations disponibles sur les pages mentionnées dans les résultats demande un décodage que les élèves ne savent généralement pas faire (repérage de l'adresse, du type de serveur, du type de page, de la date de mise à jour).

Une capacité indéniable à trouver

Nonobstant toutes ces difficultés, il faut reconnaître que les élèves ont montré une bonne capacité à trouver des documents utiles. En deux séances de travail, les élèves du groupe FPS ont visité 12 sites en moyenne et en ont retenu un peu plus de 3 pour la rédaction de leur document final. Le nombre de sites consultés au cours d'une séance de 1h30 se situe aux alentours de 6. De plus, aucun élève n'a achevé sa recherche sans avoir trouvé au moins un document intéressant.

Tout d'abord, les élèves ont su tirer parti d'une des caractéristiques essentielles des moteurs : une présentation des résultats qui classe souvent un site important parmi les premiers affichés. Ainsi un élève du groupe MBTS effectuant une recherche sur le gaspillage d'énergie a appelé Altavista avec le seul mot "gaspillage" et a pu consulter 5 sites parmi les 10 premiers affichés et même en retenir un pour la rédaction de son document final.

Ensuite, le choix des thèmes de recherche avait été effectué après avoir vérifié que les informations utiles étaient présentes sur la toile. Il convient de noter à cet égard une abondance de pages traitant de sujets d'actualité comme le nucléaire, l'environnement, ou l'énergie, tandis que les sujets classiques de l'enseignement (les connaissances de référence de la discipline comme la mécanique classique ou l'optique géométrique) sont beaucoup moins présents.

D'autre part, la coopération entre élèves a fonctionné de manière efficace tout au long des séances de recherche. Même si leurs thèmes sont différents, les élèves échangent volontiers les adresses et les sites généraux qui intéressent leurs pairs. Nous avons même constaté une méthode de travail coopérative assez efficace entre trois élèves, l'un d'eux appelant un moteur pendant que deux autres consultaient les pages mentionnées dans les résultats.

4. Discussion et propositions

Ces constatations ne diffèrent pas sensiblement des résultats d'une observation d'élèves de collège pratiquant des activités documentaires sur des questions scientifiques (Wallace, 1997). Elles nous conduisent aux propositions qui suivent.

Une représentation naïve de la toile et de son contenu

Il apparaît assez nettement que les élèves ne possèdent pas une représentation suffisamment élaborée de la toile et de son contenu pour y développer des stratégies de recherche efficaces.

En premier lieu, les élèves se comportent comme s'ils pouvaient trouver toutes les réponses attendues sur la toile. D'où une certaine déception à l'utilisation des outils généraux de recherche qui ne répondent pas à leurs attentes.

En second lieu, l'identification des auteurs des pages qu'ils visitent ou des institutions qui hébergent les serveurs qu'ils explorent ne leur est pas familière. En particulier, la distinction entre les principaux modes d'expression, institutionnelle, scientifique ou personnelle, ne leur apparaît pas clairement.

Sur ces questions, il est vraisemblable qu'une "visite guidée" de ce que peut révéler Internet en matières de ressources scientifiques et de la manière dont ces informations sont présentées, constituerait un préalable utile à toute recherche d'information.

Des outils généraux peu efficaces pour une phase d'initiation

A l'examen de la manière dont les élèves les utilisent, il ne semble pas que les outils généraux de recherche d'information, annuaires et moteurs, soient bien adaptés à une activité documentaire du moins dans une phase d'initiation. Au delà des difficultés concernant la formulation des requêtes et l'interprétation des résultats, les élèves manifestent des difficultés à planifier correctement leur activité, oubliant trop souvent de noter des références, ou omettant d'explorer des pages indiquées dans les résultats. Ces questions ont déjà été soulevées à propos de la recherche d'information dans un hypertexte (Rouet, 1998).

De ce point de vue, l'absence d'outils de mémorisation du travail effectué, l'impossibilité de rechercher facilement un mot dans plusieurs documents, alliées à la quantité importante de bruit engendré par les moteurs constituent autant d'obstacles difficiles à surmonter sans une méthode de travail rigoureuse.

Une approche plus synthétique de la recherche d'information peut être mise en œuvre avec un métamoteur comme Copernic (<http://www.copernic.com>) qui accumule les résultats et permet la mémorisation des pages ainsi qu'un affinage de la recherche. Nos premiers essais d'utilisation ont montré que ce genre d'outils dont la complexité est un peu plus importante que les moteurs habituels demande un apprentissage.

Nous étudions actuellement une solution intermédiaire en développant un outil de navigation et de recherche d'information qui comporte les principales fonctionnalités évoquées ci-dessus et qui offre en même temps un soutien technique et méthodologique à l'élève qui l'utilise.

Une formation nécessaire

Bien que l'initiation à la recherche documentaire fasse partie des programmes du collège, les élèves de lycée ne maîtrisent pas suffisamment les techniques documentaires pour être capables de les adapter d'eux-mêmes aux conditions nouvelles de recherche sur internet. Comme le souligne P. Bernahrd (Bernahrd, 1998), la maîtrise de l'information est une des clés de réussite dans une société d'information et de savoir.

Des contenus et des ressources destinés à cette formation commencent à être produits à l'instar de ce que propose le site Fromanet (<http://pages.infinit.net/formanet/formanet.html>). Une formation plus spécialisée aux techniques applicables aux documents scientifiques serait probablement nécessaire.

5. Conclusion

Une observation de plusieurs groupes d'élèves en situation de recherche d'information en sciences nous a permis de constater que ceux-ci avaient une connaissance très naïve de la toile et une maîtrise très imparfaite des outils généraux, annuaires et moteurs de recherche.

A ces outils peu adaptés, nous suggérons de substituer, au moins dans une phase d'initiation, des outils de navigation et de recherche qui facilitent la planification et le contrôle de la recherche. Ces outils actuellement en cours de développement pourraient intégrer des fonctions de conseils fondées sur une analyse de l'activité de l'élève.

6. Références

Bernahrd, P. (1998). Apprendre à «maîtriser» l'information : des habiletés indispensables dans une «société du savoir», *Education et Francophonie*, 26, 1.

Rouet, J.-F. & Tricot, A. (1998). Chercher de l'information dans un hypertexte: vers un modèle des processus cognitifs., In A. Tricot & J.-F. Rouet (eds), Les hypermédias, approches cognitives et ergonomiques. (57-74). Paris : Hermès.

Wallace, R., Kupperman, J. (1997). On-Line Search in the Science Classroom: Benefits and Possibilities, *American Educational Research Association Meeting*, Chicago, IL.

Remerciements

Cet travail a été effectué avec la collaboration de Olivier Kempf, Bernard Richoux, Monique Schwob et Martial Tarizzo.