

# LA RECHERCHE D'INFORMATIONS SUR INTERNET PAR DES LYCEENS

Analyse et assistance à l'apprentissage

**François-Marie Blondel**

INRP - Département Technologies Nouvelles et Education, 91, rue Gabriel Péri - 92120 Montrouge

Francois-Marie.Blondel@inrp.fr

**Résumé** : La formation à la recherche d'informations sur internet pose la question des outils proposés aux internautes ordinaires et incite à étudier des solutions plus adaptées. Notre projet vise à étudier la manière dont les lycéens et les étudiants s'approprient les outils de recherche et à concevoir des dispositifs capables de les assister dans cette activité. L'observation de lycéens sur des tâches contrôlées montre que ceux-ci apprécient cette activité et sont capables de trouver des informations qu'ils estiment pertinentes. L'analyse de leurs pratiques montre de réelles difficultés à formuler des interrogations efficaces et à en exploiter méthodiquement les résultats. L'assistance que nous étudions comprend actuellement un historique structuré, élaboré à partir d'une analyse de la trace et un outil d'extraction et de structuration de fragments.

**Mots-clés** : recherche d'informations, analyse des usages, suivi de l'activité de l'apprenant, utilisabilité, ergonomie des interfaces de recherche, outils d'assistance.

**Abstract** : Training in web-based information seeking raises questions about usability of general search tools available for end users and induces the study of specific solutions. The purpose of our project is to analyse how students make use of web search tools and to design specific additional help tools. The analysis of web searching behaviour of high school students shows that they enjoy this activity and that they are able to find information they consider as relevant. But they encounter real difficulties in expressing efficient queries and dealing with the results of their queries. We present two help tools : a history mechanism, built on an analysis of the main navigation events, and a tool that allows fragments extraction and organization, and may improve information processing.

**Keywords** : information retrieval, web search behaviour, usability, user interface, help tools

## INTRODUCTION

La recherche d'informations dans les bases documentaires, dans les grandes bases de textes ou sur internet, est une tâche complexe, qui met en jeu des savoirs et des savoir-faire de natures diverses qui commencent seulement à être identifiés. Cette activité est l'objet de nombreuses modélisations tant du point de vue des sciences de l'information que de celui de la psychologie cognitive (Tricot et al., 1998 ; Rouet & Tricot, 1998 ; Kolmayer, 1998) .

Les comportements de recherche d'informations sur internet ont tout d'abord été étudiés chez des utilisateurs ordinaires et experts. Ceux des lycéens et des étudiants sont beaucoup moins connus et leurs difficultés ont été assez peu analysées. Les observations de lycéens effectuées par R. Fidel (1999) ont montré que ceux-ci conduisent leurs recherches avec rapidité et souplesse et qu'ils sont satisfaits de leurs résultats mais les difficultés rencontrées plaident pour une formation spécifique et des outils plus appropriés.

Il est probable que l'usage de la toile dans des activités de recherche documentaire ira croissant dans les années à venir. La mise en place de travaux personnels encadrés dans les lycées développera certainement ces usages. De plus, la formation des étudiants à la maîtrise de l'information qui semble

jouer un rôle dans le processus d'affiliation et de réussite universitaire (Coulon, 1996) sera amenée à se développer au sein des formations méthodologiques de premier cycle.

Afin de concevoir des formations spécifiques et des outils plus adaptés à ces usages, nous nous sommes interrogés sur ce qui caractérise les activités de recherche d'informations en situation d'apprentissage et sur les principaux obstacles rencontrés par les apprenants. Sur ce point, les outils permettant une observation fine restent encore à définir.

Le projet auquel nous travaillons vise à identifier les caractéristiques de la recherche d'informations chez les lycéens et les étudiants en formation, à analyser la manière dont ils s'approprient les outils de recherche et à concevoir des nouveaux outils capables de les assister dans cette activité.

Dans cet article, nous présenterons d'abord les résultats de nos observations, puis des éléments de réflexion pour la conception d'outils et enfin une proposition d'outils complémentaires concernant notamment l'historique de la recherche et l'extraction d'informations.

## **OBSERVATION DU COMPORTEMENT DE RECHERCHE D'INFORMATIONS DES LYCEENS**

### **Méthode**

A des fins d'observation, nous avons organisé plusieurs séances de recherche d'informations sur internet pour des élèves de lycées. Durant ces séances, deux types d'observations ont été effectuées : une observation directe de la classe et des élèves permettant de noter les conditions de travail, les événements importants et les demandes émanant des individus, et un traçage de la navigation. L'observation directe a été effectuée par une personne extérieure à la classe, intervenant uniquement à la demande des élèves. Dans certains cas, cette observation a été complétée par des enregistrements vidéo ou de courts entretiens d'élèves.

#### *Traçage de la navigation*

Les études portant sur un grand nombre d'utilisateurs reposent le plus souvent sur les données de trace disponibles sur les serveurs. Ces données ne portent que sur les requêtes effectuées et permettent difficilement de suivre un utilisateur particulier.

Le logiciel de traçage que nous avons développé, s'installe sur le poste de l'utilisateur et fonctionne en relation avec le navigateur (Internet Explorer). Il peut intercepter toutes les requêtes partant du navigateur et les principales actions de l'utilisateur et notamment les clics de souris. Ceci permet une observation assez fine des interactions entre l'utilisateur et le document affiché.

Tous ces événements sont enregistrés au fur et à mesure. Ils ont ensuite été analysés pour en extraire les caractéristiques principales comme les appels de moteurs ou d'annuaire, les requêtes formulées, les sites visités et les liens suivis. Cette analyse de la trace a permis de mesurer les durées consacrées à telle ou telle partie de l'activité, fournissant ainsi des indications sur le comportement de l'utilisateur.

#### *Caractéristiques des classes étudiées*

Les observations ont concerné principalement des élèves de sciences ou en classe de sciences<sup>1</sup>. Le travail demandé aux élèves consistait à rechercher des informations sur une question posée par l'enseignant ou discutée collectivement, et à produire un document de synthèse sur cette question.

Les classes et les élèves observés sont les suivants :

---

<sup>1</sup> Toutes les données issues de ces observations n'ont pas encore été dépouillées. Des résultats complémentaires sont disponibles sur le site *Activités Documentaires et Internet en Physique et Chimie*, rubrique *Exemples* (<http://www.inrp.fr/Tecne/adipc/adipc-index.htm>).

- Une classe de première scientifique d'un lycée de centre ville sur des questions liées aux filières énergétiques de production d'électricité en 1999, durant 2 séances de 1h30 en avril 99 (PS99). Les élèves y ont travaillé par groupe de deux le plus souvent.
- Deux paires d'élèves de seconde d'un lycée de la banlieue parisienne sur une question relative à l'atmosphère, durant 3 séances de 1h en mai 2000 (S00).

Dans leur grande majorité, les participants n'étaient pas des usagers d'internet mais possédaient une pratique minimale des outils de bureautique. Des informations sur le maniement du navigateur et sur les outils de recherche d'informations leur ont été fournies en début de travail.

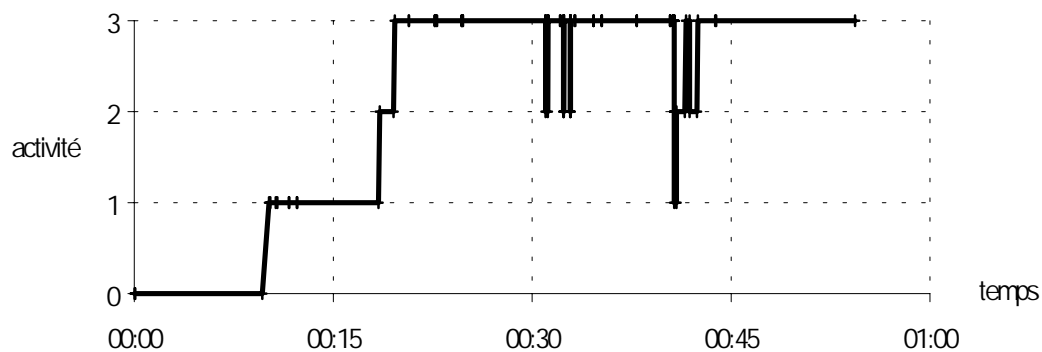
## Résultats

Nous allons présenter ces résultats en fonction des différentes tâches qui composent une recherche d'informations. Nous distinguerons : la formulation des requêtes, la consultation des résultats et le traitement des informations recueillies. Auparavant, nous donnons quelques indications sur la répartition du temps entre ces différentes tâches.

### *La répartition du temps entre les différentes tâches de la recherche*

Dans des conditions de travail imposées (durée et question), il est intéressant d'observer comment les lycéens répartissent leur temps entre les différentes composantes de leur recherche. Avec les données dont nous disposons, nous pouvons distinguer : la préparation, la formulation d'une requête, la lecture des résultats de la requête et la consultation de sites.

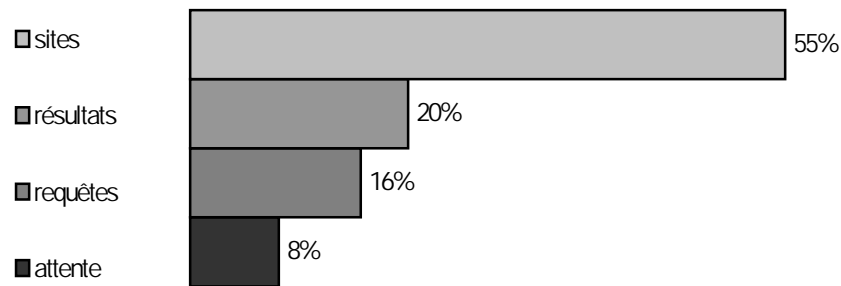
La représentation graphique de la figure 1 donne une vision chronologique de l'activité d'un utilisateur. On observe une période de préparation de 10 min en début de session (niveau d'activité codé 0, avant le premier appel d'un annuaire, codé 1). La formulation de la première requête (code 1) correspond à un temps assez important : 8 min. La lecture des résultats est assez brève (code 2) et aboutit à la consultation du premier site (code 3), assez longue (12 min) ; puis les élèves retournent à la page de résultats, consultent un autre site (code 3). Au bout de 40 min environ, ils formulent rapidement une autre requête (code 1) et consultent immédiatement deux sites. La période de 10 min en fin de session (la dernière page visitée reste affichée) est consacrée à une discussion entre les élèves.



**Figure 1.** Exemple de répartition temporelle des différentes tâches sur une séance pour un groupe (classe PS99 groupe éolienne séance n°1)

Une analyse globale des durées recueillies sur 6 groupes de PS99 montre qu'un peu plus de la moitié du temps est consacré à la consultation des sites (55%). La formulation des requêtes (16%) et la consultation des résultats rendus par les moteurs ou les annuaires (20 %) occupe un peu plus du tiers de leur temps (36% au total). Le reste correspond à des temps morts y compris les délais de réception des résultats ou des documents (8 %), délais peu importants dans ce cas.

Mais nous avons pu noter une grande variabilité entre les individus ; ainsi, le temps de formulation des requêtes (moyenne 16%) varie-t-il entre 3% et 25% du temps total. Cette variabilité se retrouve à toutes les étapes du travail.



**Figure 2.** Durées moyennes des groupes de PS99 sur les deux séances

Dans cette situation (scolaire) de recherche, les élèves se distinguent nettement des utilisateurs ordinaires dont on sait qu'ils consacrent en moyenne peu de temps à chaque recherche. Disposant d'une certaine liberté d'action, les élèves ont été confrontés à une difficulté inhabituelle, celle de gérer leur temps disponible, ici sur deux séances de travail, ce que traduisent les grandes variations dans leur répartition du temps. Des outils capables de les assister dans cet aspect de gestion de la tâche seraient certainement appréciables. En pratique, la mesure des différentes durées peut être utilisée pour conseiller l'apprenant sur la gestion de son temps.

#### *La formulation des requêtes*

Comme le montre l'exemple de la figure 1, une séance commence très souvent par l'appel d'un moteur ou d'un annuaire. Le mode d'interrogation préféré des lycéens est de loin celui de la recherche par mots clés. Il importe donc d'analyser en détail comment ils formulent leurs requêtes.

#### *Les termes des requêtes*

Les 6 groupes PS99 ont formulé 52 requêtes distinctes comportant au total 101 mots ou expressions, soit 2 termes en moyenne par requête.

Ces données sont comparables avec les formulations d'utilisateurs ordinaires : 2.35 mots par requête en moyenne pour les utilisateurs d'AltaVista (Silverstein et al., 1998), 2.4 pour ceux de Excite (Wolfram, 2000). La comparaison entre experts et novices effectuée par Hölscher et al. (1999) sur les utilisateurs du moteur allemand Fireball indique par ailleurs : 1.66 mots par requête pour les novices et 3.64 pour les experts.

L'attitude des lycéens paraît pour le moins minimaliste ; leur tendance générale est à utiliser peu de mots. 42% des requêtes de la classe PS99 ne comportent qu'un seul terme et 38% deux termes seulement.

Ainsi, pour rechercher des informations sur la filière de production d'énergie par la voie éolienne, un groupe n'a utilisé que 3 termes, employés séparément (éolienne, vent et wind), estimant avoir récolté avec ces requêtes suffisamment de documents. D'autre part, le groupe de PS99 qui a formulé le plus grand nombre de requêtes (soit 17 au total dont 13 distinctes) n'a employé que 7 termes ou expressions différents (voir tableau 1).

En règle générale, les termes des requêtes sont directement puisés dans le texte de la question. Les tentatives effectuées pour enrichir la requête avec des mots qui soient plus discriminants, sont assez rares.

Les formulations plus développées, comme celles du Tableau 1, sont plus des combinaisons des mots de la question qu'une investigation des termes potentiellement efficaces pour la recherche.

Les solutions proposées pour assister les apprenants sont le plus souvent conceptuelles : dictionnaires, thésaurus et encyclopédies. Mais les outils disponibles en ligne sont encore assez rares.

+centrale+charbon
+centrale+combustible
+centrale+pétrole
"centrale électrique"
"centrale électrique"
+centrale+fuel
"centrale électrique"+fuel
"centrale électrique"+pétrole
centrale électrique à charbon
"centrale électrique"+charbon
"centrale électrique" +"charbon"
"centrale électrique" +"charbon"+"combustible"
"charbon"+"combustible"+"centrale électrique"

**Tableau 1.** Requêtes formulées par le groupe PS99 cherchant sur la filière combustible

### *L'emploi des opérateurs*

En revanche, comme on peut l'observer sur l'exemple du tableau 1, les élèves utilisent assez volontiers les opérateurs pour formuler leurs requêtes.

Les 6 groupes PS99 ont assez largement utilisé les opérateurs qui leur avaient été présentés comme acceptables dans une requête simple. L'opérateur + (présence obligatoire du mot qui suit) se retrouve dans 16 requêtes sur 52 (31%), et l'opérateur "" (expression exacte) dans 11 requêtes sur 52 (21%). En revanche, l'opérateur - (absence obligatoire du mot qui suit) n'est jamais utilisé. D'autre part, très peu utilisent le mode recherche avancée proposé sur presque tous les moteurs.

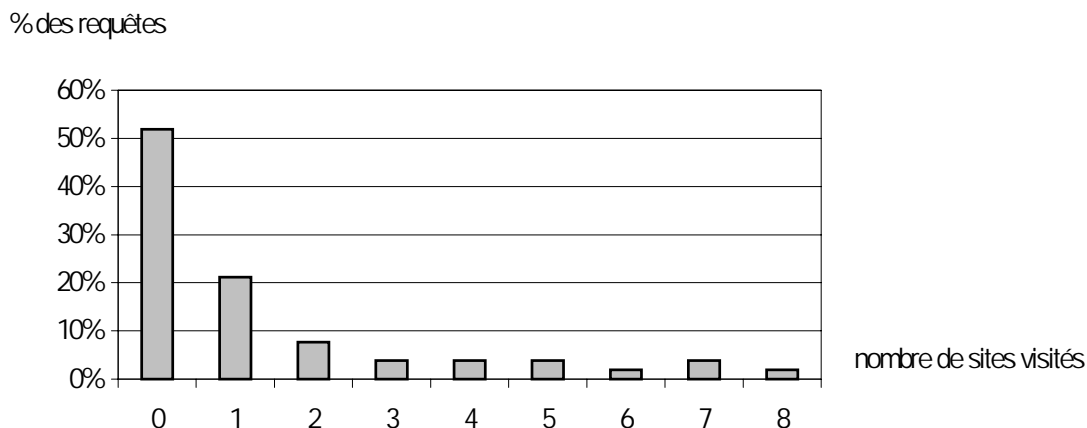
De ce point de vue, les lycéens se situent entre les experts et les novices. D'après les observations de Hölscher et al. (1999), la fréquence d'emploi de l'opérateur + est de 29% chez les experts, et de 25% chez les novices, et celle de l'opérateur "", de 25% chez les experts contre 9% chez les novices. A contrario, les experts emploient beaucoup plus fréquemment les modes de recherche avancés (34% des requêtes comportent l'opérateur AND chez les experts contre 2% chez les novices). On peut rappeler aussi que 80% des requêtes adressées à AltaVista sont sans opérateur (Silverstein et al., 1998).

Cette apparition des opérateurs dans les requêtes est rarement immédiate ; elle est plutôt utilisée par les lycéens comme un moyen destiné à en améliorer l'efficacité, lorsqu'ils jugent les résultats peu satisfaisants. Cependant, les résultats des requêtes plus complexes ne sont pas pour autant plus satisfaisants. D'après Johnson et al. (1998), une initiation préalable à l'usage des opérateurs booléens n'accroît pas de manière significative la satisfaction des lycéens. L'apprentissage du (bon) emploi de ces opérateurs pourrait plutôt passer par une analyse de leurs effets dans la pratique. Encore faut-il que les apprenants aient les moyens de réfléchir sur leurs actions et leurs conséquences.

D'autre part, on pourrait penser qu'une formulation plus complexe permet d'obtenir des résultats plus pertinents. Mais il n'est pas avéré que la formulation de requêtes complexes, c'est-à-dire avec les opérateurs appropriés, ait beaucoup plus de chances de fournir des résultats significatifs que des formulations simples. Que l'on emploie les mêmes termes dans une requête simple ou dans une requête complexe, les résultats des moteurs sont souvent très proches.

### *La consultation des résultats*

La consultation des résultats retournés par un moteur ou un annuaire aboutit à une consultation de sites ou à la formulation d'une nouvelle requête, ou encore à un abandon. Cette opération de sélection des documents à consulter est une part non négligeable de l'activité de recherche.



**Figure 3.** Distribution du nombre de requêtes en fonction du nombre de sites visités (total 52 requêtes, 71 sites dont 4 sites visités deux fois)

Les élèves de PS99 y consacrent 20% de leur temps. Pour 85% des requêtes, ils se contentent des premiers résultats, c'est-à-dire de la première page retournée par le serveur. La page des résultats suivants n'est consultée que dans 10% des requêtes, et les résultats qui suivent dans 5% des cas seulement. Ce comportement est semblable à celui des utilisateurs "ordinaires" du serveur AltaVista, qui se contentent de la première page de résultats dans 85% de leurs requêtes (Silverstein et al., 1998).

Si l'on cherche à évaluer la satisfaction des lycéens par les consultations qui découlent des requêtes, celle-ci n'apparaît pas très élevée. Le nombre de sites visités par requête est en moyenne de 1.2 pour les groupes de PS99 et le nombre de requêtes qui ne sont pas suivies par la consultation d'un site est de l'ordre de la moitié (Figure 3).

Les traces recueillies ne nous permettent pas de savoir comment les élèves "lisent" les listes de résultats rendus par les moteurs et encore moins quels critères leur permettent de sélectionner tel ou tel site pour le consulter. Sur les principaux moteurs, la présentation des résultats ne facilite pas une vision synthétique et ne fournit pas beaucoup d'éléments de structuration.

Peu satisfaits des résultats qui leurs sont rendus, certains élèves essaient différentes techniques pour améliorer leurs requêtes. Un groupe de S00 a ainsi effectué plusieurs requêtes successives sans consulter aucun site, cherchant uniquement à réduire le nombre total de résultats affiché par le moteur Voilà, passant de 2873 pour la première à 25 pour la dernière.

La présentation des résultats dans un cadre uniforme et mieux structuré permettrait aux utilisateurs de mieux percevoir les conséquences de leurs choix. Des conseils pourraient alors leur être fournis en fonction de ces choix.

#### *Consultation et utilisation des documents visités*

La consultation des pages, c'est-à-dire des documents provenant des sites visités reste la partie essentielle du travail de recherche d'informations. Pour les lycées de PS99, cette part de l'activité est prépondérante puisqu'ils y consacrent en moyenne plus de la moitié de leur temps de recherche.

Ce résultat est conforme à ce que Byrne et al. (1999) rapporte à propos d'utilisateurs expérimentés. En effet, ces derniers consacrent la majeure partie de leur temps à la tâche définie sous le terme "utilisation de l'information", et à l'intérieur de cette tâche, la lecture occupe une place largement dominante.

La navigation dans les sites à partir de l'adresse renvoyée par le moteur ou l'annuaire est le plus souvent réduite à quelques pages, voire une seule. Bien que nous n'ayons pas de données d'observation précises sur la "lecture" de ces pages, les lycéens ne semblent pas employer de méthodes systématiques pour inspecter ces documents et décider de leur intérêt pour la construction de la "réponse".

Dans les documents longs, la recherche d'indices relatifs à la question pourrait être améliorée par une indexation automatique des documents consultés et par une recherche dans ces index en comparaison avec la question.

Le nombre de documents utilisés par les lycéens pour la rédaction de leur document de synthèse est relativement limité. Dans leur ensemble, les 6 groupes de PS99 ont utilisé des pages provenant de 14 sites distincts soit en moyenne un peu plus de 2 sites utilisés par document. Tout se passe comme si les lycéens se contentaient d'un seul document par réponse. Il est d'ailleurs assez facile de vérifier qu'ils ne font pas de comparaisons entre documents différents pour répondre à une question.

## LES OUTILS DE RECHERCHE ET LA FORMATION

A l'issue de ces observations, on pourrait se poser la question de savoir si les outils *ordinaires* de recherche d'informations sur internet, annuaires et moteurs généralistes, sont particulièrement adaptés à l'initiation et à la formation des lycéens. Certes, ils sont conçus pour satisfaire un large public, généralement novice comme les lycéens ou les étudiants. Mais ils s'adressent aussi à un public plus spécialisé, comme en témoignent la présence d'options de recherche et de modes avancés.

### Outils de recherche au fonctionnement opaque

Plusieurs aspects du fonctionnement des moteurs de recherche sont à prendre en considération, comme le modèle de représentation (vectoriel, booléen), ou le calcul et l'utilisation de la pertinence.

Sur ce dernier point, les techniques de calcul de pertinence adoptées par les principaux moteurs pour retourner en tête de leurs résultats les pages les plus susceptibles d'intéresser l'utilisateur, en intégrant des heuristiques sophistiquées pour contrer les tentatives de contournement de la part des auteurs de sites, rendent la compréhension de leur fonctionnement assez opaque. A tel point que certains estiment "qu'il devient quasiment impossible aujourd'hui de se forger un modèle mental réaliste du fonctionnement des moteurs de recherche" (Bourdoncle, 1999).

A contrario, il faut observer que les utilisateurs, même novices comme les lycéens, trouvent dans les premiers résultats affichés des sites et des documents qui répondent en partie à leurs questions, et qu'ils s'avèrent capables d'améliorer, même modestement, l'efficacité de leurs requêtes. L'emploi des outils généralistes pour la formation reste donc possible, mais il devrait être accompagné d'une mise en situation et d'outils qui en facilitent la mise en œuvre (Fidel et al., 1999).

A cela, on peut ajouter un intérêt grandissant pour des outils de recherche généralistes dont les méthodes sont plus explicites et aussi plus efficaces, à l'instar de Google ([www.google.com](http://www.google.com)) qui utilise un critère de popularité, la citation, pour effectuer un classement des documents, et qui, de plus, bénéficie d'une interface très épurée et simple d'emploi.

### Interfaces d'interrogation et de visualisation

Les interfaces d'interrogation et de visualisation des résultats de moteurs sont généralement peu précises et incomplètes. Ainsi, des options implicites de recherche utiles à la formulation et à la compréhension des résultats (casse, accents, mots vides, langue, opérateurs) ne sont pas toujours affichées. Ces défauts qui handicapent les utilisateurs novices ont été mis en évidence par Schneiderman et al. (1997) qui ont analysé les interfaces de recherche dans les bases de textes et proposent quelques règles pour en améliorer l'usage : offrir une information significative en retour, permettre l'annulation de la dernière action, donner le contrôle à l'utilisateur, réduire la charge de mémorisation.

A défaut de pouvoir modifier les interfaces des moteurs, il est possible de proposer des formulaires d'interrogation qui explicitent les options en fonction du niveau d'expertise de l'utilisateur et éventuellement permettent de les régler (ce qui est partiellement fait dans certains modes avancés de moteurs). Il serait préférable de compléter ces interfaces d'interrogation par un affichage plus structuré des résultats, à l'instar de ce qu'offrent certains méta moteurs comme Copernic ([www.copernic.com](http://www.copernic.com)).

### **Navigation et complexité de la tâche**

Contrairement à ce que l'on observe à propos de la navigation dans les hypermédias, il ne semble pas que la désorientation soit la difficulté majeure des lycéens car, d'après les observations de Fidel et al. (1999) que nous avons confirmées, ils ne s'écartent que très rarement du sujet de leur recherche. Il semble plutôt que ce soit l'absence de vision synthétique de leur recherche qui rende le contrôle de l'activité plus délicat. Néanmoins, la chronologie des actions semble jouer un rôle crucial dans la gestion de la tâche.

La complexité de la tâche de recherche d'informations et l'absence de cadre formel précis pour la mener à bien expliquent la difficulté qu'éprouvent les utilisateurs à se remémorer leurs protocoles de recherche. En observant les comportements d'étudiants, Navarro-Prieto et al. (1999) ont constaté que ceux-ci ne parvenaient pas à reconstituer correctement leurs recherches ; en effet, ils tendent à oublier les moteurs et les requêtes qui ne les ont pas conduits à des résultats et même dans certains cas reconstituent de fausses requêtes. En cherchant à modéliser le comportement de recherche, Maglio et Barrett (1997) ont aussi observé que les experts eux-mêmes ne reproduisent qu'un petit nombre de points quand ils se remémorent leurs recherches.

Devant ces difficultés, les utilisateurs effectuent des retours en arrière sur des pages déjà visitées ou sur des résultats déjà consultés. L'existence d'un historique de la recherche qui réponde à certains critères ergonomiques (suppression des doublons, structuration hiérarchique, regroupement par "tâches") comme le suggèrent Tauscher et Greenberg (1997) pourrait faciliter ce travail.

### **PROPOSITIONS POUR ASSISTER L'UTILISATEUR EN FORMATION**

De nombreux logiciels, modules externes, assistants de navigation, ou applications spécialisées sont proposés pour faciliter la tâche de recherche d'informations.

Afin de ne pas masquer les difficultés inhérentes à l'apprentissage, l'option que nous étudions pour une initiation à la recherche d'informations consiste à rester le plus proche possible des outils "standards" afin d'autoriser une perception de leur fonctionnement et si possible de leur principe. Les outils que nous proposons viennent en accompagnement de ces outils généralistes, suivant une architecture semblable à celle des systèmes qui viennent se greffer sur une application pour délivrer des conseils.

Prenant en compte les observations présentées ci-dessus et celles effectuées par Alain Beaufils (1998 ; 2001) avec des élèves de collège et de lycée, l'assistant de navigation pour la recherche d'informations que nous proposons a pour objet d'organiser et de structurer l'activité de recherche.

En se greffant sur le navigateur, cet assistant suit à la trace et enregistre les actions de l'utilisateur et leurs résultats. Il incorpore les mêmes fonctionnalités de traçage que le logiciel utilisé pour les observations présentées plus haut.

En même temps que les actions importantes, l'assistant mémorise les documents visités afin d'en permettre l'analyse. Il classe les actions par recherche en respectant leur chronologie, afin d'en donner une vision globale à l'utilisateur et le cas échéant au tuteur-enseignant. Il facilite le traitement des documents par l'extraction et l'annotation et commente au besoin les actions de l'utilisateur. Avec cet assistant, les tâches de recherche de documents, d'évaluation des documents et d'extraction d'informations à partir de ces documents sont un peu mieux identifiées.

Cet assistant comprend actuellement trois composantes : un historique structuré, un outil d'extraction et d'annotation de documents, et un module de conseil. Les deux premières composantes, en cours de développement sont présentées ci-dessous, le module de conseil est en phase de définition.

#### **Un historique structuré par la recherche**

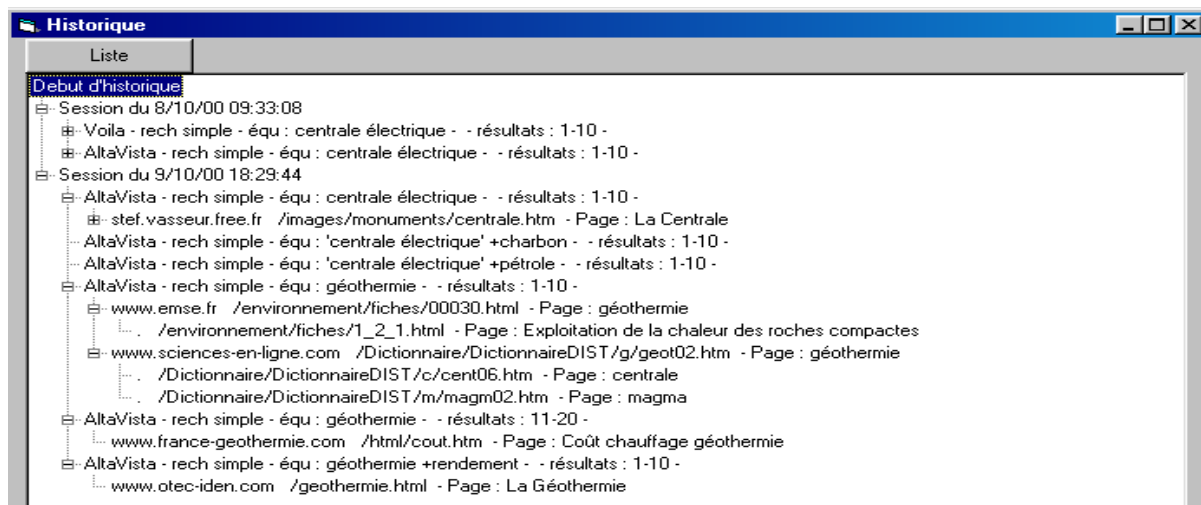
Contrairement aux historiques associés aux navigateurs, l'assistant effectue une sauvegarde automatique des documents affichés, y compris les résultats de requêtes adressées aux moteurs. Ceci



permet à l'utilisateur de retrouver, à l'identique et hors connexion, non seulement les pages mais aussi les résultats d'une recherche d'une session passée, palliant ainsi les variations possibles entre deux requêtes identiques décalées dans le temps.

Simultanément, il opère une analyse des requêtes et en dégage les caractéristiques : mots-clés, opérateurs, type de recherche, quantités de résultats affichés. Cette analyse permet d'associer les documents visités aux requêtes qui ont permis de les découvrir, ou de regrouper plusieurs requêtes pour en revoir les résultats. L'affichage conjoint d'une requête avec ses caractéristiques principales et des visites auxquelles elle a donné lieu, devrait permettre une meilleure perception des relations entre les unes et les autres.

L'analyse des requêtes est actuellement effectuée sur cinq annuaires ou moteurs généralistes : AltaVista, Google, Nomade, Voilà, Yahoo. Cette analyse peut être facilement étendue à d'autres moteurs.



*Figure 4. Historique de deux sessions présenté hiérarchiquement*

L'historique est proposé actuellement sous deux modalités d'affichage, l'une par liste chronologique ou l'autre hiérarchique organisées par sessions, recherches et sites. Cet historique est un document actif. Chaque élément, sauvegardé au cours de la navigation, peut être visualisé dans le navigateur. De plus, il peut être marqué pour en faciliter la recherche ultérieure.

L'historique est construit au fur et à mesure de la navigation ou peut être reconstruit en différé, à partir des événements de navigation précédemment tracés.

### **Un bloc-notes pour l'extraction et l'annotation**

Cet outil de prise de notes est destiné à faciliter l'extraction d'informations à partir des documents visités et la rédaction de notes associées à ces documents.

S'inspirant d'un prototype développé par Alain Beaufils (1998) pour la recherche d'informations dans un hypermédia documentaire, cet outil permet de sélectionner des pages à partir de l'historique des pages visitées, d'en visualiser le contenu, d'en extraire des fragments, d'éditer et d'annoter ces fragments, et de les structurer dans un document principal. La prise de notes fonctionne hors connexion. Le document produit est un document structuré qui peut être préparé à l'avance à partir d'un plan rédigé a priori, ou construit au fur et à mesure de la consultation. Ce document est éditable et les références aux documents d'origine sont automatiquement conservées, permettant à la fois la citation et le remords.

La recherche de documents et la prise de notes sont utilisées en alternance, le passage de la première à la seconde se faisant par l'historique qui mémorise les actions et les documents.

### Aides complémentaires

Plusieurs aides complémentaires sont en cours d'étude ou de réalisation. Parmi celles-ci, des guides d'activités qui peuvent être visualisés simultanément dans le navigateur, présentent aux apprenants le sujet de leur recherche et le travail qui leur est demandé. Ils comportent des suggestions d'outils en fonction des étapes de leur recherche (encyclopédies, moteurs et annuaires), et des modes d'emploi de ces outils dans le contexte du domaine visé.

D'autres outils d'assistance sont à l'étude : des formulaires d'interrogation de moteurs, une présentation des résultats de moteurs analysés, une indexation automatique des documents sauvegardés.

### CONCLUSION

La formation à la recherche d'informations sur internet soulève la question des outils proposés aux internautes ordinaires et incite à étudier des solutions plus adaptées à l'apprentissage.

L'observation des lycéens sur des tâches contrôlées montre que ceux-ci apprécient cette activité de recherche documentaire et sont capables de trouver des informations qui répondent à leurs questions. Leur comportement s'apparente plutôt à celui des novices mais leurs recherches présentent néanmoins des aspects plus systématiques et plus approfondis.

Malgré une aisance vite acquise dans la manipulation des logiciels de navigation et d'édition de texte, les lycéens éprouvent de réelles difficultés à formuler des interrogations efficaces et à en exploiter méthodiquement les résultats. Plus généralement, il semble que la gestion de l'ensemble de l'activité soit particulièrement difficile.

Sans masquer les déficiences réelles des outils généralistes, la présentation d'un historique élaboré à partir d'une analyse de la trace, vise à offrir une vision plus globale de l'activité au fur et à mesure qu'elle se déroule et à faciliter le contrôle de l'ensemble de la tâche par l'utilisateur. L'extraction de fragments et leur structuration par l'utilisateur vise à encourager un traitement plus détaillé des informations recueillies en vue de la production d'un document de synthèse. Des conseils méthodologiques, fournis a priori ou en fonction de l'activité devraient compléter ce dispositif d'assistance à l'apprentissage de la recherche d'informations sur la toile.

### Remerciements

Les travaux présentés ci-dessus ont bénéficié des contributions de Monique Schwob, Olivier Kempf, et Bernard Richoux, pour l'observation des lycéens, et de Jean-Claude Le Touzé et Martial Tarizzo, pour la conception et la réalisation des outils de traçage et d'assistance.

### BIBLIOGRAPHIE

- Beaufils A. (1998). "Aides à la recherche d'informations dans les environnements hypermédias", in J-F. Rouet, B. de la Passardière (Eds), *Hypermédias et Apprentissages 4*, Actes du quatrième colloque Hypermédias et Apprentissages, Poitiers, Paris : INRP-EPI, p. 73-86.
- Beaufils A. (2001). "Quelles aides pour la recherche d'informations sur internet au collège et au lycée", *ce volume*.
- Bourdoncle F. (1999). "Panorama et perspectives des outils de recherche d'informations textuelle sur Internet", in *Actes de la conférence Information, Documentation, Transfert des connaissances, IDT'99*,
- En ligne : [http://www.idt.fr/idt/pages\\_fra/actes/actes99/acte34.html](http://www.idt.fr/idt/pages_fra/actes/actes99/acte34.html), consulté le 16 mai 2001.
- Byrne, M. D., John, B. E., Wehrle, N. S., Crow, D. C. (1999). "The tangled web we wove: A taskonomy of WWW use", in *Human Factors in Computing Systems: Proceedings of CHI 99*, Reading : Addison Wesley, p. 544-551.
- Coulon A. (1996). "Penser, classer et catégoriser: l'efficacité de l'enseignement de méthodologie documentaire à l'université", *Espace universitaire*, n°15, oct. 1996,
- En ligne : <http://www.ccr.jussieu.fr/urfist/coulon.htm>, consulté le 5 avril 2001.

- Fidel R, Davies R., Douglass M., Holder J., Hopkins C., Kushner E., Miyagishima B., Toney C. (1999). "A Visit to the Information Mall: Web Searching Behavior of High School Students", *Journal of the American Society for Information Science*, vol. 50, n° 1, p.24-37.
- Hölscher C., Strube G. (1999). "Web Search Behavior of Internet Experts and Newbies", in *WWW9 proceedings*, En ligne : <http://www9.org/w9cdrom/81/81.html>, consulté le 16 mai 2001.
- Johnson B., Szabo M.(1998). "Instruction, Gender, Search Success, and Search Satisfaction on the World Wide Web: A Research Study", in *Proceedings of Edmedia/Edtelecom '98*, En ligne : <http://www.quasar.ualberta.ca/edmedia/TIES/readings/Irefjoh.html>, consulté le 16 mai 2001.
- Kolmayer E. (1998). "Démarche d'interrogation documentaire et navigation", in J-F. Rouet, B. de la Passardière (Eds), *Hypermédiats et Apprentissages 4*, Actes du quatrième colloque Hypermédiats et Apprentissages, Poitiers, Paris : INRP-EPI, p. 121-134.
- Maglio P. P., Barrett R., (1997). "How to Build Modeling Agents to Support Web Searchers", in C. Paris, C. Tasso, (Eds), *User Modeling: Proceedings of the Sixth International Conference, UM97*, New York : Springer, p. 5-16.
- Navarro-Prieto, R., Scaife, M., & Rogers, Y. (1999). "Cognitive Strategies in Web Searching", in *Proceedings of the 5th Conference on Human Factors & the Web*, juin 1999, En ligne : <http://zing.ncsl.nist.gov/hfweb/proceedings/navarro-prieto/index.html>, consulté le 12 juin 2001.
- 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 hypermédiats, approches cognitives et ergonomiques*, Paris : Hermès, p. 57-74.
- Tricot A., Pierre-Demarcy C., El Boussarghini R. (1998). "Un panorama des recherches sur l'activité mentale de l'utilisateur d'un hypermédia", in E. Bruillard, B. de la Passardière, G.-L. Baron (coord), *Le livre électronique, Sciences et Techniques Educatives*, vol. 5, n° 4, p.371-400.
- Shneidermann, B., Byrd, D., Croft, W.B., (1997). "Clarifying Search: A User-Interface Framework for Text Searches", *D-Lib Magazine*, En ligne : <http://www.dlib.org/dlib/january97/retrieval/01shneiderman.html>, consulté le 12 juin 2001.
- Silverstein, C., Henzinger M., Marais H., Moricz M. (1998). Analysis of a Very Large AltaVista Query Log. *SRC Technical Report 1998-014*, Compaq Systems Research Center, Palo-Alto, California.
- Tauscher, L., & Greenberg, S. (1997). "How people revisit web pages: Empirical findings and implications for the design of history systems". *Int. Journal of Human-Computer Studies*, vol. 47, p. 97-137.
- Wolfram D., (2000). "A Query-Level Examination of End User Searching Behaviour on the Excite Search Engine", in *Proceedings of ACSI 2000*, En ligne : <http://www.slis.ualberta.ca/cais2000/wolfram.htm>, consulté le 16 mai 2001.