

## AIDE À LA RECHERCHE D'INFORMATIONS SUR INTERNET, AU COLLÈGE ET AU LYCÉE

**Alain BEAUFILS**

INRP (Tecne) - 91, rue Gabriel Péri - 92 120 MONTROUGE  
alain.beaufils@inrp.fr

**Résumé :** Nous présentons dans cet article les résultats d'une étude exploratoire menée sur quelques élèves de collège et de lycée ayant pour tâche de répondre à une question à thème large en utilisant uniquement des informations trouvées sur Internet. Nous nous référons au modèle cognitif de Rouet et Tricot pour décrire les difficultés rencontrées par ces élèves au cours des différentes étapes de leur recherche d'informations (RI) : l'évaluation de la question, la recherche des sources, l'identification des informations pertinentes et leur traitement pour construire la réponse. Nous centrons plus particulièrement nos observations sur les problèmes liés à l'utilisation d'Internet (construction de requêtes, repérage des références adéquates dans les résultats fournis par un moteur). Si certaines difficultés semblent pouvoir s'atténuer par un entraînement régulier ou par l'application de méthodes adaptées, d'autres doivent trouver une solution en utilisant des outils spécifiques. Nous nous inspirons d'un prototype développé au cours d'une étude précédente pour définir quelques caractéristiques d'un dispositif d'aide à la RI sur Internet.

**Mots-clés :** recherche d'informations, Internet, aide, logiciel éducatif, restructuration d'informations

**Abstract :** This paper analyses web searching behavior of middle school and high school students through field observation in class. Two groups of students were asked a complex question ; they had to search the web for answers and to write down their findings in a structured document. The difficulties and problems they encountered concern : goal evaluation, document searching, identification of relevant information and production of a structured synthesis. Some difficulties are more specific of web searching, like building queries that match search engines' requirements and selecting relevant references among the results given by search engines. If some difficulties may be reduced by suitable training, others should be reduced by help tools. Inspired by our previous prototype for hypermedia searching, we define the functionalities of a help tool that may reduce some identified difficulties.

**Keywords :** information retrieval, Internet, educational software, help, information restructuring.

## **INTRODUCTION**

La quantité d'informations dont peuvent disposer aujourd'hui les élèves de tous niveaux scolaires a fortement augmenté et n'a cessé de se valoriser grâce à une présentation multimédia de qualité croissante. Les supports se sont diversifiés (vidéo, CD-ROM, DVD-ROM ou Internet) tandis que les moyens de recherche et d'accès font de plus en plus appel aux nouvelles technologies, en particulier à l'informatique. Les activités de recherche documentaire se développent largement à l'école, au collège ou au lycée, vivement encouragées voire imposées par des directives institutionnelles. Par exemple, la mise en place dans les lycées des travaux personnels encadrés (TPE).

Le document est devenu un objet d'étude (Ferrandon, 1999) et la recherche d'informations est aujourd'hui une discipline transversale qui doit s'acquérir dès l'école primaire (Dinet, Rouet & Passerault, 1998). Or la recherche d'informations est une activité complexe qui met en jeu des processus cognitifs particuliers bien décrits par différents chercheurs, notamment par Rouet et Tricot qui en ont élaboré un modèle (Rouet & Tricot, 1998).

Depuis plusieurs années, nous nous intéressons aux problèmes de l'activité documentaire dans le secteur scolaire. Au cours de précédentes études basées sur l'observation de situations pédagogiques diverses (Beaufils, 1998a et 1998b), nous avons cherché à décrire les difficultés que rencontrent habituellement des élèves de différents niveaux lorsqu'ils cherchent des informations dans des bases documentaires de type hypermédia. Pour réduire certaines de ces difficultés, nous avons proposé des solutions méthodologiques ou instrumentales. Dans cet article, nous décrivons quelques difficultés cognitives ou techniques plus spécifiquement liées à la recherche d'informations sur Internet. Nous nous demandons notamment si les consignes et les outils que nous avons proposés pour faciliter l'exploitation de bases documentaires restreintes pourraient s'adapter au fonds documentaire plus complexe que constitue Internet.

## **PRÉSENTATION GÉNÉRALE DE L'ÉTUDE**

À quelques mois d'intervalle, nous avons conduit dans deux établissements scolaires de la banlieue parisienne (collège et lycée), deux observations complémentaires sur quelques élèves (trois élèves de quatrième et huit élèves de seconde). Dans les deux situations, les élèves étaient volontaires et ont accepté de traiter une question portant sur un thème relativement large en utilisant uniquement des informations trouvées sur Internet. Leur recherche était limitée aux seuls documents rédigés en langue française. L'activité s'est déroulée au CDI et comprenait une séance d'entraînement suivie de plusieurs sessions hebdomadaires d'environ une heure chacune (quatre sessions pour les élèves de quatrième, six sessions pour les élèves de seconde).

Les élèves engagés dans ce travail sont peu nombreux, de niveaux différents et répondent à des questions différentes. Il ne s'agit donc que d'une étude observatoire dont le seul objectif est de repérer des comportements et des processus qui paraissent se répéter dans certaines situations : par exemple, la manière dont les

élèves s'approprient des méthodes et des outils pour chercher des documents et les exploiter. Les résultats dont nous faisons état ici n'ont qu'une valeur relative et sont à considérer comme des pistes de réflexion propres à faire évoluer ces méthodes et ces outils ou à en imaginer de nouveaux qui pourront ensuite être testés au cours d'études plus approfondies.

## MÉTHODE

### Les questions posées aux élèves

Les questions suivantes ont été posées aux différents groupes :

1. aux élèves de quatrième : *L'enfance victime de la violence des adultes. Sa représentation en littérature, dans la chanson et au cinéma (donner quelques exemples).*

2. aux élèves de seconde (filles) : *Les efforts faits par les femmes pour améliorer leur statut et leur image vous paraissent-ils avoir eu des résultats positifs dans tous les domaines de la vie ?*

3. aux élèves de seconde (garçons) : *Quels sont les effets et les dangers à court et à long terme des différentes substances absorbées par les sportifs ? Qu'en pensez-vous ?*

Ces questions abordent volontairement des thèmes plutôt larges et sont d'une certaine complexité : plusieurs concepts s'articulent et plusieurs domaines se croisent. Répondre à ces questions demande un certain temps de préparation (compréhension de l'énoncé et choix de stratégies diversifiées). Plusieurs explorations successives sont nécessaires avant de disposer d'un volume d'informations suffisant. La construction de la réponse oblige à un travail de synthèse et de restructuration.

### Le modèle cognitif

Pour conduire nos observations, nous nous référons au modèle cognitif de Rouet & Tricot (1998). D'après ce modèle, les processus mis en œuvre au cours d'une recherche d'informations se déroulent selon trois étapes fondamentales qui se répètent de façon cyclique tout au long de l'activité jusqu'à ce que le but de la recherche soit atteint. Ces étapes sont 1) l'évaluation du besoin d'informations ressenti par le sujet pour résoudre le problème ou répondre à la question, 2) la sélection, dans l'environnement exploré, des sources susceptibles de contenir les informations recherchées, 3) le traitement des informations trouvées qui, combinées les unes aux autres et intégrées aux connaissances que le sujet possède déjà sur le domaine, servent à construire la solution du problème ou la réponse à la question.

Nous tenons compte du fait qu'Internet présente des différences importantes avec les bases documentaires sur CD-ROM que nous avons utilisées au cours des observations précédentes : les documents y sont beaucoup plus nombreux, non validés et aucune description générale des contenus, tels que table des matières ou index, n'est directement consultable. Pour trouver des informations dans cette masse de documents, il est nécessaire de recourir à des outils de recherche spécifiques et

souvent complexes. En fait, la recherche d'informations (RI) sur Internet donne lieu à deux processus qui se succèdent : dans un premier temps, une recherche documentaire aboutissant à trouver les références de documents supposés pertinents et dans un deuxième temps, la consultation de ces documents dans le but d'y trouver des informations pertinentes (Dinet, Rouet & Passerault, 1998).

À côté des modèles cognitifs théoriques proposés par les chercheurs, des pédagogues comme Yves Léveillé ou Hélène Guertin (Groupe Formanet)<sup>1</sup> proposent des méthodes de travail basées sur l'observation et sur la pratique pour guider la RI sur Internet. Si l'on compare ces méthodes entre elles, des similitudes apparaissent, notamment dans la manière de planifier l'activité. Dans la plupart des cas, on retrouve les trois étapes fondamentales modélisées par Rouet et Tricot. Les conseils et les consignes que nous avons donnés aux élèves en début d'activité s'apparentent à ces méthodes.

### Les outils et les consignes

Chaque élève a utilisé deux dispositifs d'accompagnement distincts. D'une part un dossier individuel de *recherche de documents* créé à l'aide du métamoteur *Copernic*, d'autre part un dossier individuel de *prise de notes* créé à l'aide d'un logiciel de traitement de texte.

Le métamoteur *Copernic* a été utilisé pour formuler les requêtes, lancer les recherches et recueillir les résultats. Plusieurs raisons nous ont amenés à choisir ce logiciel, en particulier sa puissance et sa rapidité (plusieurs moteurs fonctionnent simultanément et une limite au nombre de documents à trouver par moteur peut être fixée à l'avance). L'ergonomie générale est claire et relativement facile à mémoriser par les élèves. Après chaque recherche, les résultats sont présentés sous forme d'une liste de références aux documents trouvés. Les élèves peuvent adapter la présentation de cette liste en fonction de leurs besoins, par exemple trier, cocher ou effacer certaines références. De plus, ils ont la possibilité de conserver tous les résultats (listes des références et documents téléchargés). En revanche, ils ne peuvent pas regrouper en une liste unique les résultats sélectionnés à la suite de recherches distinctes.

Le dossier de *prise de notes* remplissait la double fonction de guide méthodologique et d'aide mémoire. Il a été créé pour fixer les états intermédiaires du travail, favorisant à la fois la planification de la recherche et l'archivage des éléments trouvés ou produits. Il comprend deux parties : l'une réservée aux notes diverses (*indices, mots clés, plan provisoire de la réponse, projets de requêtes, sélection d'adresses*), l'autre réservée à l'élaboration de la réponse.

Dans le cadre de cette étude, il n'a pas été possible de regrouper les deux dispositifs de *recherche* et de *prise de notes* en un seul outil de travail, un peu à la manière du prototype que nous avons développé au cours d'une précédente étude et que nous avons installé dans une base documentaire hypermédia (Beaufils, 1998b)<sup>2</sup>.

---

1 Formanet - Groupe d'échange entre pédagogues, notamment sur les problèmes de recherche d'informations sur Internet (<http://www.ac-poitiers.fr/pedago/coll-lyc/formanet/sommaire.htm>).

2 La description détaillée de ce prototype peut être consultée à l'adresse suivante : <http://www.inrp.fr/Recherche/Terminees/9699/Rec40117.htm>.

Une fiche de travail comportant des conseils pour l'utilisation des outils et des consignes méthodologiques a été remise à chaque élève au début de l'activité.

### Le recueil des données

À la fin de chaque session, les traces de l'activité de chaque élève ont été recueillies à partir de plusieurs sources complémentaires : le résultat des recherches effectuées sur Internet (archives du logiciel *Copernic*), les références des documents visités (historique de la navigation) et l'état de la production écrite (copie du fichier individuel de prise de notes).

## QUELQUES RÉSULTATS

### La préparation de l'activité (évaluation du besoin d'informations)

Les élèves devaient effectuer plusieurs exercices préparatoires avant de se livrer à la recherche d'informations proprement dite : lire attentivement la question, rechercher aussi largement que possible les mots clés pouvant définir et préciser le thème de la question (élargissement du champ sémantique), construire un plan provisoire de la réponse et élaborer quelques formules de recherche à partir des mots-clés dégagés. Le résultat de ces travaux devait être porté dans le dossier individuel de prise de notes.

#### *Identification du besoin d'informations, planification de l'activité*

Dans l'ensemble, il apparaît que les élèves des deux groupes comprennent correctement les questions qui leur sont posées et identifient assez précisément les thèmes auxquels ces questions se rattachent. Toutefois, selon les questions et le niveau des élèves, les exercices préparatoires se révèlent inégalement fructueux.

Les élèves de quatrième ne parviennent pas à développer de champs sémantiques au-delà des mots de l'énoncé. Tout au plus réussissent-ils à leur associer quelques synonymes ou quelques mots proches. Malgré la consigne donnée, le temps passé à ce travail est bref (moins de dix minutes). Un élève consacre un peu plus de temps à l'exercice, mais les mots qu'il trouve sont des mots voisins qui n'élargissent pas sensiblement le champ sémantique. Les plans de réponse restent succincts et reproduisent la structure de l'énoncé (tableau 1).

<b>mots et axes sémantiques définis dans l'énoncé</b>	<b>notes en vrac</b>	<b>exemple de plan provisoire</b>
<i>axe 1</i> : la violence	violence, viol battu, maltraité, insultes	texte qui définit le sujet : l'enfance victime de la violence
<i>axe 2</i> : les adultes (les auteurs)	adultes	
<i>axe 3</i> : les enfants (les victimes)	enfants, enfance, infantile	
<i>axe 4</i> : la littérature, le cinéma, la chanson (la représentation)	littérature, théâtre cinéma, film chanson, chants, chanteurs	représentation littéraire représentation au cinéma représentation dans la chanson
		solutions (associations)

Tableau 1. Notes en vrac et plan provisoire de la réponse (élèves de 4<sup>e</sup>).

Les élèves de seconde ont une meilleure connaissance initiale du thème de leur question. Ils s'acquittent avec une certaine maîtrise des exercices préparatoires qui se prolongent pendant une session complète (une heure) - voire un peu plus pour certains élèves - sans que des signes de lassitude soient détectés. Les notes sont le plus souvent abondantes, variées et pertinentes. Les mots-clés repérés dans l'énoncé font émerger de nombreux mots nouveaux qui enrichissent le champ sémantique de la question. En revanche, certains aspects de la question ne sont pas pris en compte, notamment ceux qui font appel à une appréciation personnelle. Dans l'ensemble, les plans provisoires construits à partir des notes reprennent assez exactement la structure générale de la question mais chaque section du plan est relativement bien détaillée (tableau 2).

mots de l'énoncé	notes en vrac réparties selon les axes sémantiques développés par les élèves	exemple de plan provisoire
		1. Introduction
effort	<i>histoire des luttes</i>	2. Les différentes luttes de la femme droit de vote, avortement accouchement sous X
statut	<i>droits acquis, à acquérir (lois)</i>	3. Les statuts de la femme
	<u>droits généraux</u> parité homme/femme : droit de vote, travail/salaire, expression, sécurité, sport	grandes femmes lois en faveur des femmes, parité
	<u>droits spécifiques</u> Avortement accouchement sous X	
image	<i>Images de la femme</i> <del>image traditionnelle (maternité, foyer) *</del> image artistique, image commerciale médias, télévision, magazines(mode)	4. Les différentes images de la femme art, télévision, magazines
résultats + ou -		
avis personnel	<del>moyens d'améliorer le statut *</del>	5. Conclusion

(\*) Les expressions barrées correspondent à des notes que les élèves ont inscrites puis effacées par la suite.

*Tableau 2. Notes en vrac et plan provisoire de la réponse (élèves de seconde).*

*Des champs sémantiques ont été développés autour de trois axes repérés dans l'énoncé.*

*En revanche, deux axes (résultat des luttes et évaluation personnelle) ont été peu pris en compte.*

### **Évolution des notes et des plans**

Dans les deux situations observées, on remarque assez généralement que les notes préparatoires portées dans le dossier individuel au début de l'activité sont rarement complétées ou réactualisées par la suite, notamment lorsque des *événements* se produisent au cours de la consultation. Par exemple, lorsque de nouveaux mots-clés sont découverts, les élèves les utilisent aussitôt pour construire de nouvelles formules de recherche dans le logiciel *Copernic* mais ils ne les ajoutent pas à la liste des mots-clés qu'ils ont élaborée dans leurs notes. De la même manière, lorsque les élèves trouvent des extraits d'informations contenant des éléments

nouveaux par rapport à leur plan, ces extraits sont directement ajoutés aux extraits déjà sélectionnés pour la construction de la réponse et la structure initiale du plan n'est pas modifiée.

### ***Éléments de discussion à propos de l'activité d'évaluation***

La richesse des champs sémantiques développés autour d'un thème est liée à l'abondance et à la diversité des connaissances initiales que les élèves possèdent sur ce thème et qu'ils réussissent à activer. Le domaine étudié et la maturité des élèves sont sans doute à prendre en compte mais il semble possible d'améliorer le processus d'activation par un entraînement régulier et par le recours à des dispositifs d'aide spécifiques. Par exemple l'utilisation – à condition qu'ils existent – de *thesaurus* généraux conçus pour l'usage scolaire. À défaut, certains enseignants conseillent de consulter des sommaires d'encyclopédies ou les *renvois* proposés par les dictionnaires (synonymes, contraires, analogues).

Par ailleurs, les quelques remarques suivantes peuvent aider à expliquer le fait que le travail préparatoire ne soit pas suffisamment pris en compte et entretenu pendant toute la durée de l'activité :

- Lorsque les élèves découvrent de nouvelles informations, leur attention est principalement occupée à évaluer l'intérêt de ces informations. Certaines tâches peuvent alors être oubliées ou négligées en raison d'une saturation de la mémoire de travail.
- Si la partie du dossier dans laquelle sont insérées les notes préparatoires n'est pas suffisamment visible et accessible, la mise à jour de ces notes risque de ne pas s'effectuer correctement. Les élèves ont des difficultés à maintenir une activité régulière et simultanée dans plusieurs zones de travail distinctes.
- Malgré les consignes données, les élèves ont tendance à considérer que l'évaluation de leur besoin d'information sous forme de notes écrites est un travail de début d'activité sur lequel ils ne voient pas nécessairement l'utilité de revenir par la suite.

Un dispositif ergonomique permettant à la fois d'assurer la disponibilité permanente de l'espace réservé aux notes et une assistance à leur mise à jour éviterait peut-être la dispersion de l'attention des élèves et inciterait ceux-ci à mieux utiliser cette partie de leur travail.

### **La recherche et la sélection des sources**

Dans les requêtes formulées à l'aide du métamoteur *Copernic*, les termes utilisés doivent être juxtaposés et séparés par un espace auquel l'utilisateur choisit d'attribuer la valeur *et* ou la valeur *ou* (la valeur *et* est attribuée par défaut). Il est également possible de demander la recherche d'une expression exacte à condition de placer cette expression entre guillemets. C'est la configuration par défaut (*espace = et*) qui a été la plus spontanément utilisée.

### **Formules élaborées par les élèves et résultats obtenus**

À partir de leurs notes préparatoires, les élèves de seconde élaborent des requêtes assez bien diversifiées. La plupart des formules qu'ils utilisent sont construites de façon un peu systématique et combinent généralement deux termes : le mot-clé principal de la question (*femme* ou *dopage*) associé successivement aux différents mots clés qui définissent les sections de leur plan. Par exemple [*« droit de vote » femme*], [*avortement femme*], [*dopage loi*] ou [*dopage argent*]. Selon les élèves, le nombre de requêtes formulées au cours des six sessions est compris entre deux et dix. Les documents pertinents récoltés sont nombreux et facilement identifiés dans les résultats fournis par les moteurs.

Les élèves de quatrième utilisent eux aussi leurs notes préparatoires pour construire leurs requêtes dans lesquelles ils croisent également de façon systématique les différents mots-clés dont ils disposent. Ces requêtes sont relativement nombreuses - de huit à dix-sept par élève au cours des quatre sessions -, elles sont cohérentes et en rapport avec le sujet traité mais peu diversifiées car elles n'utilisent que peu de mots différents. Trois types de requêtes ont été construites :

- des requêtes à deux termes visant à obtenir des informations sur le thème général de la violence envers l'enfance, par exemple [*violence enfance*], [*enfant maltraité*] ou [*violence infantile*].
- des requêtes à trois termes visant à obtenir des informations plus spécifiquement centrées sur l'un des trois domaines demandés, par exemple [*littérature violence enfant*], [*enfant maltraité cinéma*] ou [*chanson violence enfance*].
- quelques requêtes n'utilisant qu'un seul terme généralement trouvé en cours de consultation.

Le tableau 3 fait apparaître la rentabilité médiocre de ces requêtes. Le nombre de recherches effectuées et le nombre de références récoltées témoignent de la persévérance, voire de l'obstination des élèves. Si le nombre de références se rapportant au thème général de la *violence envers l'enfance* atteint 43% des documents trouvés, en revanche le nombre de références ayant un rapport plus précis avec le thème de la question est faible (5% des documents trouvés).

Nombre de requêtes	Nombre de termes utilisés par requête	Documents trouvés	Documents pertinents
3	2 termes [ <i>violence + enfance</i> ]	95	41 [43%]
28	3 termes [ <i>violence + enfance + domaine</i> ]	457	23 [5%]
5	1 terme (divers)	127	1 [0,8%]
36	total	679	65 [9,6%]

Tableau 3. *Résultats globaux obtenus par les différents types de requêtes (élèves de 4e).*

### **Repérage des sources pertinentes dans les listes de résultats**

Le logiciel *Copernic* affiche les résultats de chaque recherche sous forme d'une liste de références et fournit quelques informations sur chacun des documents trouvés, notamment quelques mots en caractères gras servant de titre, l'adresse (URL) agrémentée de quelques lignes de texte pouvant servir à identifier le contenu.

Indépendamment de leur pertinence apparente par rapport au sujet, nous avons évalué la lisibilité de toutes les références récoltées par les élèves de quatrième en distinguant les références clairement identifiables de celles qui pouvaient prêter à confusion. Nous avons ensuite confronté la sélection faite par les élèves à notre évaluation. Lorsque les références sont clairement identifiables, les documents potentiellement pertinents sont à peu près bien repérés. En revanche, un nombre important de références que nous avons estimées difficilement identifiables sont conservées par les élèves (62 % de leur sélection). Dans presque tous les cas, les expressions rencontrées dans ces références peuvent évoquer le thème de la question par la présence de certains mots comme *guerre*, *colère*, *alcoolisme*, *menaces* ou *enfance abusée*. D'autres références font allusion à l'un des domaines demandés. Par exemple *bibliographie*, *filmographie*, *livres pour la jeunesse* ou *chansons et enfants*. On peut attribuer la sélection de ces références à une mauvaise interprétation du texte, limitée à la seule prise en compte d'unités lexicales correspondant au sujet traité (Caron, 1989 ; Fayol et al, 1992 ; Dinet, Rouet & Passerault, 1998).

Le tableau 4 montre les différences existant entre les deux situations observées.

	Nb de requêtes	Réf. trouvées	Réf. sélectionnées	rentabilité
Élèves de 4 <sup>e</sup>	8 à 17	226 [22]	29 [11]	5%
Élèves de seconde	2 à 10	201 [>100]	25 [21]	10%

Tableau 4. *Références sélectionnées par les élèves (résultats moyens par élève). Les références adéquates sont notées entre crochets.*

Les élèves de quatrième lancent des recherches qui s'avèrent peu rentables, ce qui les incite à renouveler l'exercice mais sans varier notablement le contenu de leurs requêtes. Finalement, le nombre de références qu'ils parviennent à sélectionner est faible et de plus comporte 60% de références inadéquates. Les requêtes formulées par les élèves de seconde ont une meilleure rentabilité, ce qui explique leur nombre plus limité. Au total, ils sélectionnent un peu moins de références que les élèves de quatrième mais la proportion de références adéquates est nettement plus élevée.

#### **Consultation des documents : accès par URL et par navigation**

Dans les deux situations observées, la consultation des documents se fait soit à partir des URL sélectionnées dans les résultats, soit en suivant les liens proposés dans les documents visités (Tableau 5).

La plupart des élèves ne consultent pas tous les documents dont ils ont sélectionné les références (environ un tiers de ces documents ne sont pas affichés). Quant aux documents consultés à partir de liens, ils représentent 36% de la consultation totale chez les élèves de seconde. Les élèves de quatrième utilisent davantage ce procédé (58% des accès) mais sans beaucoup de succès. Ils espèrent découvrir par ce moyen des informations qu'ils n'ont pas trouvées dans les pages affichées à partir des URL. Quelques élèves se livrent à l'exploration systématique de sites entiers (par exemple des sites contenant des textes de chansons). La démarche utilisée dans ce cas est infructueuse et s'apparente au feuilletage.

	Références sélectionnées	Consultation par URL	Consultation par liens	Total consultation
Élèves de 4 <sup>e</sup>	29 [11]	20 [7]	28 [2]	48 [9]
Élèves de seconde	25 [21]	16 [15]	11 [6]	27 [21]

Tableau 5. Documents consultés à partir des URL sélectionnées ou des liens rencontrés.  
Les nombres indiquent les résultats moyens par élève.  
Les références adéquates sont notées entre crochets.

### Éléments de discussion à propos de la recherche et de la sélection des sources

Internet n'est pas une encyclopédie et les documents qu'on y trouve ne traitent pas forcément de tous les sujets. Une consultation régulière de la toile permettrait sans doute aux élèves de se faire une représentation plus précise du type de documents qu'elle contient et d'évaluer quelle type de recherche il convient d'y mener. Par exemple les documents correspondant à la question posée aux élèves de 4<sup>ème</sup> sont sans doute peu nombreux. Par ailleurs, la précision insuffisante des requêtes formulées par les élèves, la limite des moteurs de recherche et la médiocre lisibilité des résultats sont autant de causes qui ont encore diminué les chances de les repérer.

La recherche d'informations sur Internet demande des compétences d'experts en particulier l'utilisation des moteurs de recherche. Les élèves éprouvent certaines difficultés à construire des requêtes précises avec des termes reliés entre eux par des opérateurs booléens (*et, ou, sauf*) délicats à manier par des non-spécialistes. Par exemple, l'articulation des trois concepts contenus dans la question posée aux élèves de quatrième - [*la violence*] envers [*l'enfance*] dans le domaine de [*la littérature*] – a un sens précis qu'il est difficile de traduire dans une requête. Les formules élaborées par les élèves expriment ces concepts mais n'en définissent pas l'articulation car les termes sont simplement juxtaposés. Ces termes par ailleurs très généraux provoquent la récolte de nombreux documents indésirables contenant les mots sans pour autant correspondre au sujet.

Les listes de résultats fournies par les moteurs ne donnent qu'une idée approximative du contenu réel des documents récoltés. Dans le cas de thèmes généraux comme le *statut des femmes* ou les *effets du dopage* pour lesquels il est aisé de récolter de nombreux documents pertinents, la simple consultation de ces listes peut paraître suffisante. Pour des questions plus complexes comme *la violence envers l'enfance dans la littérature*, d'autres méthodes permettant de consulter les résultats de façon plus précise doivent être trouvées. La constitution d'une base de documents téléchargés paraît une solution intéressante.

- Les analyses automatiques de texte (éventuellement des systèmes d'indexation) permettent d'identifier le contenu réel des documents. Il est également possible d'effectuer des recherches personnalisées dans une base de textes téléchargés.
- Les documents téléchargés sont stables et rapidement disponibles, ce qui permet de les consulter plusieurs fois et de les comparer entre eux. Il est possible, à partir de ces documents, de générer des listes de références plus faciles à consulter que celles fournies par les moteurs.

Les élèves ont des difficultés à exploiter correctement les listes de références obtenues après chaque. De nombreux documents sont ignorés en raison d'une difficulté à gérer des listes distinctes. Il semble souhaitable que les élèves disposent en

permanence d'une seule liste de références. Cette liste peut être créée (si possible de façon automatique) à partir des résultats de recherche, de l'historique de la consultation ou d'une base de documents téléchargés. Une liste unique de références a pour fonction d'aider les élèves à mieux *planifier* et à mieux *contrôler* leur travail. Elle permet notamment de retrouver facilement les références des documents déjà consultés et de garder une trace de ceux qui ne le sont pas encore. Elle peut permettre également de classer les documents en fonction des besoins individuels. Divers remaniements doivent être possibles : ajouts, retraits, déplacements de références. Cette liste est active et permet d'accéder directement à tous les documents qu'elle cite.

## Le traitement des informations trouvées

### Identification de l'information dans les documents sélectionnés

Le Figure 1 montre que le nombre de documents exploités dans les synthèses finales est relativement faible par rapport au nombre de documents pertinents sélectionnés et consultés.

Chez les élèves de quatrième, le nombre de documents exploités est peu élevé (de trois à huit documents par élève). Dans l'ensemble, la pertinence des extraits prélevés est satisfaisante. Dans certains documents, l'information n'a pas été trouvée. Plusieurs explications peuvent être avancées :

- l'information n'existe pas dans le document. L'évaluation des résultats fournis par les moteurs ne s'est pas faite correctement, certaines informations ont pu prêter à confusion.
- l'information existe dans le document mais l'élève ne réussit pas à l'identifier soit parce qu'elle est exprimée dans un langage complexe ou spécialisé (par exemple un rapport scientifique ou juridique), soit parce qu'elle se trouve noyée dans un contexte opaque que l'élève ne réussit pas à déchiffrer de façon précise, soit parce que la manière dont elle est présentée (vocabulaire, typographie) est différente de celle à laquelle l'élève s'attendait.

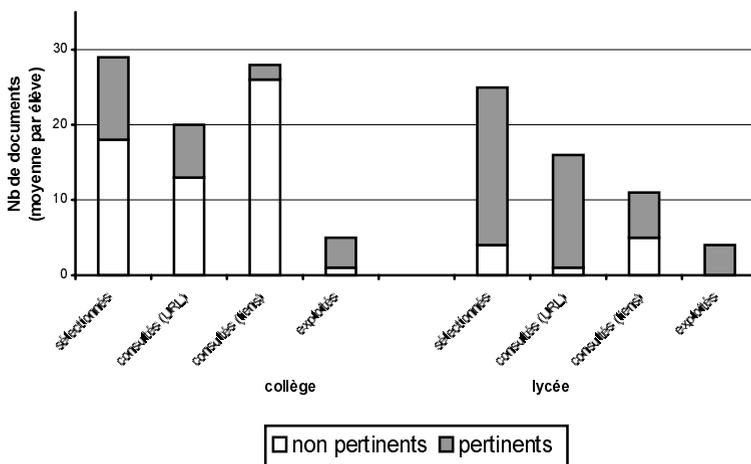


Figure 1. Documents consultés à partir des URL sélectionnées ou des liens rencontrés au cours de la consultation (résultats moyens par élève).

D'une manière générale, les élèves de seconde n'ont pas eu de difficultés à identifier les informations pertinentes présentes dans les documents qu'ils ont consultés. Au contraire, plusieurs informations leur ont semblé redondantes et les ont amenés à faire des choix. Aucune tentative de synthèse entre plusieurs documents complémentaires n'a été observée.

Les élèves des deux groupes ont tendance à considérer qu'un document est valide lorsqu'il contient l'information cherchée. Cette confiance excessive est évidemment liée au fait que les élèves sont peu entraînés à mettre en doute les sources d'informations qui leur sont proposées dans le cadre des exercices scolaires (manuels, ouvrages du CDI).

Il convient sans doute d'entraîner les élèves à identifier la nature des sites qu'ils consultent (examen de l'adresse, recherche de la page d'accueil ou du sommaire dans lesquels on trouve généralement des informations sur le site, leurs auteurs et les buts qu'ils recherchent).

### ***Traitement des extraits***

Les activités de traitement proprement dites – production de synthèses ou de résumés, citations intégrées ou non à une production personnelle – ne sont que rarement observées. On assiste plus fréquemment au prélèvement intégral d'extraits généralement assez larges. Dans ce cas l'opération s'apparente davantage à une nouvelle sélection de documents qu'à une démarche de traitement.

La section *documents* dans laquelle les élèves de seconde étaient invités à reporter leurs extraits bruts avant de les intégrer dans leur synthèse s'est avérée d'une efficacité relative. Les extraits utilisés restent copieux mais nous assistons à plusieurs reprises à des tentatives de morcellement de ces extraits et à leur insertion dans des sections différentes du plan.

### ***Structuration de la réponse et évolution du plan initial***

Chez tous les élèves, on note le souci de ne pas s'écarter du plan défini au début de l'activité. Dans la plupart des cas, aucune nouvelle section n'apparaît. Tout au plus, quelques sous sections sont créées (événements). Dans les cas observés, les élèves cherchent toujours à intégrer ces informations nouvelles dans une section existante du plan même si l'adéquation n'est pas évidente. Un autre phénomène plusieurs fois constaté est l'insertion dans une section du plan d'un large extrait lui-même structuré en sections. Certains élèves adoptent alors la nouvelle structure en surcharge ou en remplacement de celle qu'ils ont élaborée.

Le tableau 6 présente l'évolution du plan d'un élève de seconde entre le début et la fin de l'activité.

Plan initial	requêtes	doc. exploités		plan final
		extraits sélectionnés	extraits utilisés	
2. les pays touchés France ; Amérique ; Europe de l'Est	[France dopage] [pays +dopage]	doc 4 doc 8	doc 4	le dopage dans différents pays <i>doc 4. Le dopage en Italie</i>
1. Le dopage dans le cyclisme ; culturisme ; tour de France ; sponsor			doc 8	2. les sports les plus touchés <b>le cyclisme</b> <i>doc 8. Expertise dopage et sport (ext 1)</i>
3. les substances leur origine ; guerre mondiale ; fabrication	<b>[EPO amphétamine anabolisant]</b> [dopage médecin] [évolution dopage]	doc 1,2, 3  doc 7		3. les substances <b>a) leur but</b> <i>doc 8. Expertise dopage et sport (ext 2)</i> <b>b) en général</b> <i>doc 8. Expertise dopage et sport (ext 3)</i> <b>c) les contrôles</b> <i>doc 8. Expertise dopage et sport (ext 4)</i> <b>d) les effets secondaires, parfois mortels</b> <i>doc 8. Expertise dopage et sport (ext 5)</i>
4. l'argent illicites ; sponsors	[sponsors dopage] [dopage argent] [« effet secondaire » dopage]	doc 6	doc 6	4. l'argent  <i>doc 6. Exploits sportifs et dopage</i>
5. les causes performance ; argent ; triche	[guignol dopage]	doc 5		5. les causes  <i>doc 8. Expertise dopage et sport (ext 6)</i>
	9	8	3	

Tableau 6. *Évolution du plan (élève de seconde).*

La structure du plan d'origine est maintenue à quelques inversions près. Le découpage est un peu confus et parfois déséquilibré. Certaines sections sont redondantes (4 et 5), d'autres mal définies et un peu pléthoriques (3). Si le thème général est à peu près traité, l'essentiel de la question est escamoté : effets et dangers du dopage, avis personnel sur le sujet.

Sur huit documents sélectionnés, trois seulement sont utilisés (pas nécessairement les plus appropriés). L'un des documents sert de source principale d'informations puisque six extraits sont répartis dans les différentes parties de la synthèse finale.

Inversement, certains élèves ont des difficultés à alimenter toutes les catégories de leur plan initial. Certains élèves font appel à leurs connaissances personnelles pour rédiger certains chapitres. Aucun élève des deux groupes ne traite les sections de son plan pour lesquelles il était prévu d'exprimer un avis personnel.

### **Éléments de discussion à propos du traitement de l'information**

La section *documents* du dossier individuel prévue pour stocker de longs extraits de documents en attente d'un traitement plus raffiné apparaît comme un dispositif à la fois complémentaire d'une base de documents téléchargés et

intéressant à perfectionner. Son utilisation peut constituer une étape de traitement préparatoire.

Quant à un outil de restructuration assistée, il faut le concevoir comme un dispositif utilisant une série de cellules à capacité limitée (de type Post-it™ électroniques) dans lesquelles l'élève est amené à insérer peu de texte. En revanche le nombre de ces cellules n'est pas limité. L'élève est ainsi contraint de morceler l'information en *grains* élémentaires qui favorisent par la suite une restructuration personnelle. La position de ces cellules est facilement modifiable et catégorisable. On peut de surcroît – comme c'est le cas dans notre prototype – masquer partiellement le contenu des cellules pour mieux appréhender la structure du plan.

## CONCLUSION

Dans les deux situations observées, on retrouve à chaque étape du travail des difficultés comparables à celles qui ont été observées lors de l'utilisation de bases documentaires sur CD-ROM : l'évaluation détaillée du domaine de la question, le repérage des références adéquates, l'identification des informations pertinentes dans un contexte souvent volumineux et l'intégration d'extraits dans un plan de réponse structuré.

À ces difficultés s'en ajoutent d'autres plus spécifiques à Internet comme l'utilisation d'un moteur ou d'un métamoteur et la construction de requêtes interprétables par cet outil.

Certaines de ces difficultés peuvent être atténuées par un entraînement régulier à la RI et par l'utilisation de méthodes comme celles que nous avons proposées dans cette étude. Nous pensons toutefois que ces méthodes ont des limites et que la réussite de certaines opérations n'est pas uniquement liée à la simple application de consignes : par exemple, développer un champ sémantique, formuler une requête ou élaborer une synthèse.

En revanche, des aides instrumentales déjà testées dans des situations de RI peuvent être adaptées à une activité sur Internet. Un dispositif inspiré de notre précédent prototype est actuellement à l'étude. Ce dispositif est prévu pour accompagner les élèves tout au long de leur activité, sa fonction principale étant de soulager la mémoire de travail. Grâce à son interface, plusieurs types d'actions peuvent être menés simultanément (prises de notes, formulation de requêtes, recherche de documents et gestion des références). La construction d'une base documentaire à partir des documents téléchargés complète ce dispositif. Certaines interactions automatiques entre les diverses fonctions de l'outil peuvent être imaginées : la mise à jour des listes (mots-clés ou références), l'existence de liens actifs entre les références et la base documentaire peut aider les élèves à mieux planifier leur travail. Enfin une section de stockage temporaire d'extraits de documents associée à un système de traitement structuré en cellules souples doit pouvoir aider les élèves à améliorer la construction de leur réponse.

**BIBLIOGRAPHIE**

- Beaufils A. (1998b). « Aide à la recherche d'informations dans les environnements hypermédias », in J.-F. Rouet & B. de La Passardière (éds), *Hypermédias et Apprentissages 4*, Actes du 4<sup>ème</sup> colloque Hypermédias et apprentissages, Poitiers, 15, 16, 17 octobre 1998, Paris : INRP/EPI, p. 73-86.
- Beaufils A. (1998a). « Aide à l'exploitation de bases hypermédias », in A. Tricot & J.-F. Rouet (éds), *Les hypermédias, approches cognitives et ergonomiques*, Paris : Hermès, Hypertexte et hypermédias, p. 191-209.
- Caron J. (1989). *Précis de psycholinguistique*. Paris : Presses Universitaires de France.
- Dinet J., Rouet J.-F. & Passerault J.-M. (1998). « Les “nouveaux outils” de recherche documentaire sont-ils compatibles avec les stratégies cognitives des élèves ? », in J.-F. Rouet & B. de La Passardière (éds), *Hypermédias et Apprentissages 4*, Actes du Quatrième Colloque Hypermédias et Apprentissages, Poitiers, 15-17 octobre 1998, Paris : INRP/EPI, p. 149-162.
- Fayol M. *et al.* (1992). *Psychologie cognitive de la lecture*, Paris : Presses Universitaires de France.
- Ferrandon M.-C. (1999). « Une offre plus lisible pour faciliter la recherche », *Les dossiers de l'ingénierie éducative*, n° 29, octobre 1999, Paris : CNDP, p. 2-6.
- Rouet J.-F. & Tricot A. (éds) (1998). « Chercher de l'information dans un hypertexte : vers un modèle des processus cognitifs », *Les hypermédias, approches cognitives et ergonomiques*, Paris : Hermès, Hypertexte et hypermédias, p. 57-74.