

INTERNET MODE D'EMPLOI

Pas de droit de cité pour le

JACQUES RIOT ^[1]

Nos enseignements évoluent, et nous demandons de plus en plus des travaux de recherche personnelle... au risque de nous trouver devant une problématique nouvelle pour nous, celle du plagiat.

Quand peut-on dire qu'il y a plagiat ? Quels problèmes pose-t-il ?

Que faire pour le prévenir ? Comment le détecter ?

Autant de questions auxquelles il nous faut tenter de répondre.

En 2007, une étude de l'université de Lyon (voir « En ligne ») montrait que 90 % des enseignants admettaient avoir été confrontés au copier-coller dans les travaux de leurs étudiants. Il faut dire que la chose est si aisée, dès lors que l'on a un accès à internet et un logiciel de traitement de texte, que résister à la tentation serait quasi surhumain. Qui d'ailleurs s'en prive, des élèves aux profs jusqu'aux plus hautes autorités ministérielles... en passant par les rédacteurs de pages Web eux-mêmes ? Et pourquoi d'ailleurs s'en priver ? Cela n'a rien de répréhensible ou d'infamant en soi... à condition toutefois de respecter certaines règles et d'éviter certains écueils.

C'est la facilité même du procédé qui pose problème. D'une part, elle est une incitation non seulement au plagiat, mais au pire des plagiats : l'acte du copier-coller ne garantit même pas, contrairement au recopiage manuel, que son « auteur » ait ne serait-ce que lu le texte original. D'autre part, comme nous l'avons vu dans l'article précédent (« Info ou intox ? », *Technologie* n° 173), elle jette le discrédit sur le contenu des sites, lequel contenu, outre qu'il n'offre pas dans bien des cas les garanties d'une publication papier, peut être lui-même issu d'un copier-coller, voire d'une chaîne de copier-coller dont l'origine se perd dans les méandres du Web. Alors, rappelons que la première des règles, pour une citation, c'est d'en donner la source, et d'en vérifier la crédibilité.

Certains plagiats sont réalisés sans intention de tricher. Cette facilité du copier-coller, son utilisation courante pour transmettre des informations à ses proches (courriels, blogs...) en font une pratique non considérée, *a priori*, comme répréhensible. D'ailleurs, en primaire et au collège, où il est souvent demandé une « recherche » internet sur tel ou tel sujet, il n'est attendu en réalité qu'une simple collecte de sources ou de citations. Un élève peut donc penser réaliser un travail de qualité, pour ce qui n'est en fait qu'un assemblage de données recueillies.

[1] Professeur de construction mécanique au lycée Jean-Macé de Vitry-sur-Seine (94).

mots-clés

informatique, internet, média, outil et méthode

Certes, le choix des sources et leur structuration en un ensemble cohérent sont un travail en soi. Comme l'affirmait, en 2002, Jean de Lire, chargé de mission à la cellule « cyberécole » de l'administration de l'Enseignement de la Communauté française en Belgique, « s'inspirer de données, qu'elles soient en ligne ou non, pour reconstruire un thème, sachant qu'on ne l'a pas inventé, c'est ça l'école » (cité par Vincent Braun dans « Copier-coller », *La Libre Belgique*, 26 juillet 2002, www.lalibre.be/article_print.phtml?art_id=72959).

Mais la naïveté des jeunes est à relativiser. À la question « Pourquoi recourez-vous au copier-coller ? », les étudiants lyonnais ont répondu : par facilité pour 59,7 % d'entre eux ; par manque de temps pour 34,8 % ; parce que « tout le monde le fait » pour 8,2 % ; parce que c'est une « pratique sans risque de sanction » et que « les professeurs ne voient pas la différence » pour 5,8 %.

Alors, n'oublions pas que la raison d'être d'une citation est d'enrichir, illustrer, expliquer ou étayer un propos. Il est donc impératif de préciser en préalable à nos apprenants ce qu'est un plagiat, notion très floue pour eux, de leur donner les attendus et les méthodes à suivre pour les travaux demandés, mais aussi de leur faire prendre conscience des bénéfices qu'ils peuvent tirer d'un travail personnel, qui seul permet de se situer dans son apprentissage, étape nécessaire pour cerner ce qui reste perfectible et progresser.

Plagiat et droit d'auteur

Dans l'Antiquité, le plagiaire était « celui qui vendait comme esclaves des hommes libres, ou faisait, pour son compte, commerce des esclaves d'autrui », et « par extension, un voleur d'enfants » (*Dictionnaire de l'Académie française*, 9^e édition). C'est aujourd'hui la « personne qui s'approprie ce qu'elle a trouvé dans les ouvrages d'autrui ou s'en inspire indûment » (*ibid.*).

Il existe diverses situations de plagiat : la copie d'un passage de livre, revue, page Web, l'utilisation d'images, graphiques, données..., la reprise à son compte d'une idée originale d'un auteur en la réécrivant, la traduction intégrale ou partielle d'un texte, tout cela sans en citer sa source, l'utilisation d'un travail produit dans un autre cours sans avoir obtenu au préalable l'accord du professeur, l'utilisation du travail d'une autre personne en le présentant comme sien, même avec l'accord de cette personne, l'achat d'un travail, sur le Web ou par d'autres moyens.

Rappelons que toute création de l'esprit est soumise au droit d'auteur de façon systématique, et donc protégée par la loi. Une création de l'esprit peut être un

plagiat !

texte, quel qu'en soit le support (livre, journal, pièce de théâtre, logiciel, site Web, etc.), une œuvre d'art (peinture, dessin, image, sculpture, photographie, architecture, etc.), un logiciel ou encore une œuvre musicale ou audiovisuelle. Sont exclus de cette définition les idées, concepts, méthodes, mots du langage courant. L'article L. 122-4 du Code de la propriété intellectuelle précise que « toute représentation ou reproduction intégrale ou partielle faite sans le consentement de l'auteur ou de ses ayants droit ou ayants cause est illicite. Il en est de même pour la traduction, l'adaptation ou la transformation, l'arrangement ou la reproduction par un art ou un procédé quelconque ».

Toutefois, l'article L. 122.5 stipule des exceptions, par exemple s'il s'agit d'une œuvre appartenant au domaine public ou si l'on fait une copie exclusivement destinée à un usage personnel. Et « lorsque l'œuvre a été divulguée, l'auteur ne peut interdire [...] la représentation ou la reproduction d'extraits d'œuvres, sous réserve des œuvres conçues à des fins pédagogiques, des partitions de musique et des œuvres réalisées pour une édition numérique de l'écrit, à des fins exclusives d'illustration dans le cadre de l'enseignement et de la recherche, à l'exclusion de toute activité ludique ou récréative, dès lors que le public auquel cette représentation ou cette reproduction est destinée est composé majoritairement d'élèves, d'étudiants, d'enseignants ou de chercheurs directement concernés, que l'utilisation de cette représentation ou cette reproduction ne donne lieu à aucune exploitation commerciale et qu'elle est compensée par une rémunération négociée sur une base ».

Dans certains cas, les étudiants pourraient arguer que la source trouvée précise le droit de la reproduire librement. C'est le cas de la « licence d'art libre », qui est une extension du *copyleft* (ou « gauche d'auteur ») – concernant les logiciels – appliquée aux textes, images, sons. Richard Stallman, programmeur et militant du logiciel libre à l'origine du projet et de la licence publique générale GNU (ou GPL), en donne une définition : « L'idée centrale du gauche d'auteur est de donner à quiconque la permission d'exécuter le programme, de le copier, de le modifier, et d'en distribuer des versions modifiées (mais pas la permission d'ajouter des restrictions [de droits] de son cru) » (« The GNU Operating System and the Free Software Movement », in DiBona, Ockman, Stone *et al.*, *Open Sources : Voices from the Revolution*, O'Reilly, 1999).

Mais la règle ne s'arrête pas à la loi, et l'intérêt d'un travail de recherche ne repose pas uniquement sur le

résultat final, mais plutôt sur les capacités développées et la démarche suivie pour y parvenir.

Limiter le recours au plagiat

Pour inciter l'élève à citer ses sources, il est nécessaire de lui fournir un mode d'emploi. Quelques règles sont incontournables :

- Les citations sans modification doivent être mises entre guillemets, l'auteur et la source indiqués soit à la suite de l'extrait et entre parenthèses soit par une note de bas de page ou en fin du document.
- Les références de la citation contiennent nécessairement son auteur, la source dont elle est extraite (publication, ouvrage, site internet...) ou son contexte pour une citation orale.
- Si l'on choisit de reformuler ou de paraphraser un auteur, cela doit être introduit par une formule du type « selon Untel », « à ce sujet, Untel estime que », « pour Untel », « le rapport Untel relève que »..., en précisant la source.
- Les illustrations doivent être, elles aussi, référencées, leur auteur et la source donnés.
- Il est bienvenu, en fin du document, de citer les ressources qui ont aidé à réaliser le travail (bibliographie, webographie).

Il s'agit en premier lieu de valoriser l'honnêteté intellectuelle et, pourquoi pas, de l'intégrer dans l'évaluation du travail lorsque celui-ci a un fond original ou personnel, et ce, même si la forme laisse à désirer. Cela peut procurer un sentiment de fierté du travail réalisé et des efforts fournis, renforcer l'estime de soi, stimuler la curiosité et le plaisir d'apprendre.

Mais il faut bien reconnaître qu'en l'absence de règles et de sanctions claires et graduelles la tentation du plagiat risque d'être forte, comme le confirment les réponses des étudiants à l'enquête lyonnaise. Ces règles pourraient être établies de façon plus officielle par un contrat de non-plagiat en début d'année, comme dans certaines universités, ou par leur intégration au règlement intérieur, par exemple.

On peut aussi tenter de limiter le recours au copier-coller par des procédés pédagogiques :

- **Des exposés oraux** suivis d'échange avec la classe et le professeur demanderont à l'élève, si la règle a bien été précisée à l'avance, la bonne maîtrise, l'appropriation du sujet, ce qui reste en définitive un des buts de l'exercice.
- **Effectuer un suivi par étapes** du travail, demander de fournir les sources trouvées et que celles-ci soient multiples, varier les travaux demandés au cours des années limitera la possibilité de plagier.

- Poser une problématique nécessitant un point de vue personnel impliquera l'élève et stimulera son désir de s'exprimer. Par exemple, la demande d'un choix de technologie pour un problème posé nécessite un état des lieux de l'existant, une comparaison des avantages et inconvénients et une argumentation.

- Inclure une demande d'interview à certains sujets imposera un travail personnel.

- Rappeler les règles et les sanctions, en demandant systématiquement d'inclure à la fin du travail rendu une phrase signée stipulant que ce travail est personnel, qu'il cite toutes ses sources et ne comporte pas de plagiat, et ce, même si un contrat en ce sens a été signé en début d'année, responsabilisera l'élève.

Nos apprenants pensent souvent que les enseignants sont moins au fait qu'eux des technologies de communication – ce qui est bien souvent la réalité. Mais rien n'empêche de leur rappeler, sans forcément les citer, que l'on dispose d'outils pour déceler les plagiat et que l'on connaît les sites proposant des travaux déjà rédigés.

Détecter un plagiat

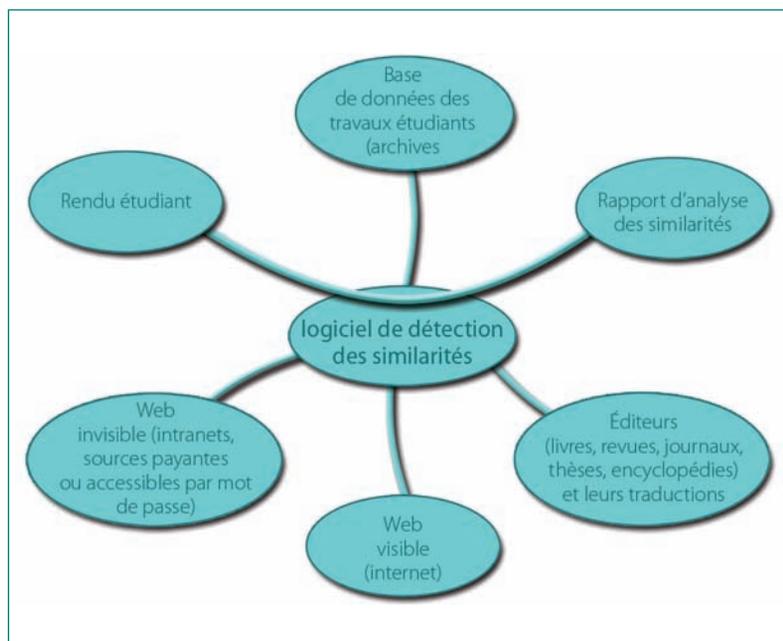
Quand peut-on s'interroger sur l'éventualité d'un plagiat ? Un travail qui détonne des productions habituelles, par son style, notamment, une abondance de références spécialisées très pertinentes, une disparité de style ou de mise en pages (diverses polices de caractères, marges non uniformes, espacements ou alignements variés, sauts de ligne incongrus...), des enchaînements peu logiques de paragraphes, une incohérence des appels de graphiques ou d'images, une sensation de patchwork portent souvent la marque du copier-coller. La présence d'hyperliens (rompus) en grand nombre en atteste l'usage, et presque à coup sûr que la source originelle est Wikipédia ; des adresses internet obsolètes indique un travail relativement ancien, donc plagié, voire autoplagié (production personnelle réutilisée).

La simplicité du copier-coller est une arme à double tranchant : elle facilite le plagiat, mais permet également de le débusquer aisément. Si le paragraphe d'un devoir, par exemple, tranche trop franchement avec le reste du texte, une fois copié et collé dans son moteur de recherche préféré, il est probable qu'on le retrouve mot pour mot dans l'un des résultats... ou dans plusieurs, auquel cas il n'y a qu'une seule et unique source – celle dont il faudrait évaluer le sérieux. Encore reste-t-il à la déterminer. Dans le meilleur des cas, qui est aussi le plus fréquent, il s'agit de Wikipédia, qui constitue cependant plus une base de recherche qu'une référence incontestable et donc suffisante. Mais beaucoup d'informations inexacts d'origine indéterminée sont allègrement dupliquées de site en site et finissent, parfois, par se retrouver dans le travail d'un élève ou étudiant. Le nombre de bouches qui colportent une rumeur n'en fait pas une vérité... C'était l'objet de mon précédent

article. Rappelons donc une autre règle : un travail de recherche doit s'appuyer nécessairement sur une analyse critique des informations récoltées.

Des outils logiciels ou des pages internet permettent d'aller plus loin, en lançant la recherche sur un nombre de mots beaucoup plus élevé, notamment. Leur principe de base est de tronçonner le texte soumis en fragments de phrases composés d'un certain nombre (qui peut être paramétrable) de mots. La recherche de similitude s'effectue sur internet et parfois dans une base de données (documents publiés ou déjà testés, rapports, thèses...) en utilisant un ou plusieurs moteurs de recherche tels que Google, Yahoo et Bing. Le rapport de synthèse, quelquefois envoyé par courriel, précise le taux d'originalité, le pourcentage de similitude, et localise les passages similaires en regard des sources concernées. Les documents acceptés sont principalement des fichiers texte (.txt, .rtf, .doc, .docx, .odt), Adobe Acrobat (PDF, pages internet) et, parfois, PowerPoint (.ppt, .pptx).

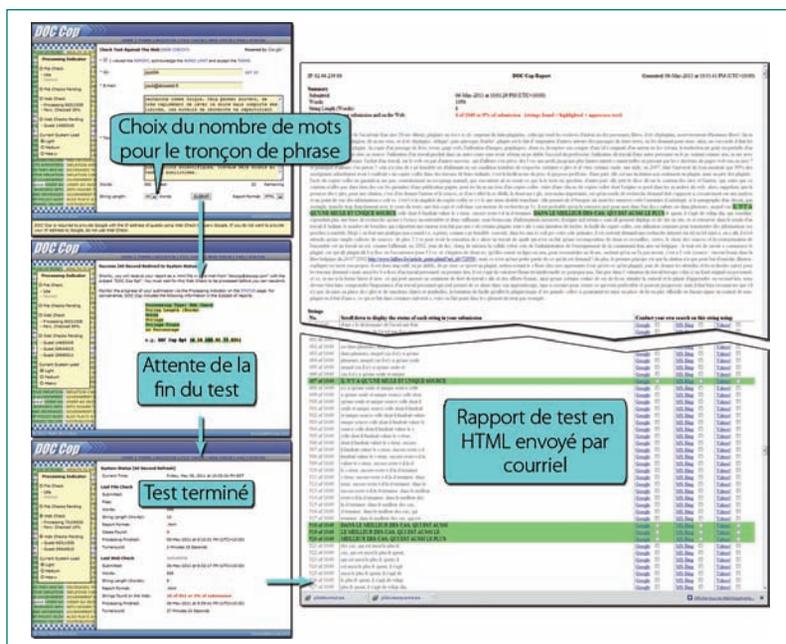
Toutefois, une partie des sources reste inaccessible ; c'est le cas de la « galaxie Gutenberg », constituée des textes non encore numérisés, mais aussi du « Web invisible » ou « Web profond », sauf en cas d'accord spécifique avec l'éditeur. Le Web invisible, source davantage pertinente, est de 100 à 500 fois plus vaste que le Web visible et croît plus rapidement que ce dernier. Il est composé des intranets (universités, écoles, entreprises, institutions, journaux...), des sources payantes ou accessibles par mot de passe (les encyclopédies, les sites offrant, moyennant finance, de faire le travail demandé à votre place comme <https://www.mesdevoirs.net>...) ainsi que des pages Web non référencées par les moteurs de recherche sur demande expresse du *webmaster* ou contenant des pages trop importantes **1**.



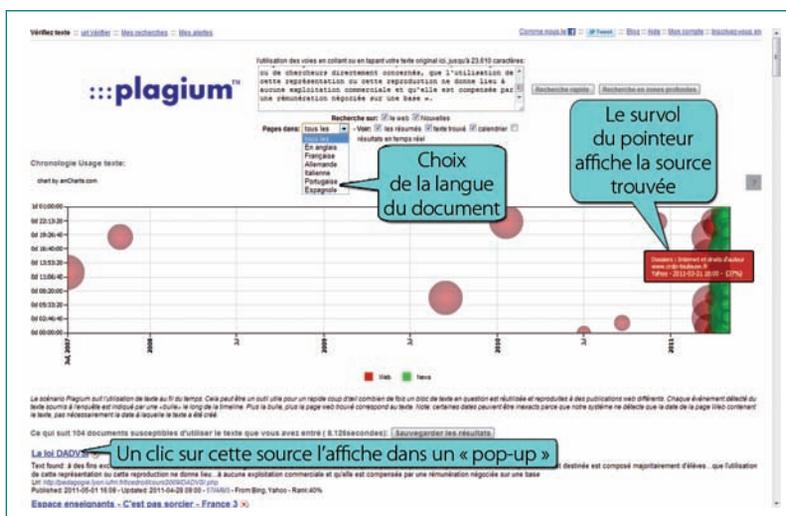
1 Les relations d'un logiciel de détection avec son environnement



2 La détection par Viper



3 La détection par Doc Cop



4 La détection par Plagium

Voici une liste de quelques outils gratuits :

- **CopyTracker** est issu d'un projet de l'École centrale de Lille en partenariat avec l'université de Genève. Il permet de comparer un fichier (.docx, .doc, .odt, .html, .txt) avec un ou plusieurs fichiers choisis, de rechercher les similarités sur des bases de données personnelles et sur internet. <http://copytracker.ec-lille.fr/cts.php?action=index>

- **Article Checker** et **Plagiarism Checker** recherchent si des extraits du texte sont présents sur le Web en utilisant les moteurs de Google et Yahoo. Le texte à soumettre doit être copié et collé dans la fenêtre du site.

www.articlechecker.com
www.plagiarismchecker.com

- **Viper** est un programme (pour Windows exclusivement) qui accepte les fichiers de tous les formats (.doc, .docx, .pdf, .html, .odt, .rtf, .txt, .s, .cs, .app, .java, .ppt, .pptx...), qu'il compare avec plus de 10 milliards de sources, sur internet, mais aussi dans ses propres bases de données. Il travaille sur la base de fragments de 5 mots au minimum de façon que l'utilisateur puisse détecter les éventuelles paraphrases. Viper affirme être le concurrent gratuit de Turnitin, un des leaders du marché (voir plus loin). L'inscription est nécessaire, l'utilisation gratuite ne limite pas le nombre d'analyses ni la taille du document. La confidentialité est assurée : votre document intégrera la base de données de Viper à des fins de comparaison, mais ne sera accessible qu'au bout de neuf mois à d'autres étudiants – et jamais avec la version payante (environ 5 \$ par mois, 10 \$ le trimestre, 15 \$ le semestre). Le rapport précise le pourcentage de parties reprises citées sans et avec guillemets, affiche les parties concernées dans le document et dans la source, et permet d'accéder à cette dernière si elle est sur internet. **2**

Pour le télécharger : <http://gateway.scanmyessay.com>

- **Doc Cop** accepte les fichiers Word (.doc, .docx) et PDF, et les analyse soit fichier contre fichier avec un minimum de 20 mots et un maximum de 50 000, soit à travers le Web (« Web Check ») avec un minimum de 20 mots et un maximum de 1 000. Il permet de choisir le nombre de mots composant le fragment de phrase à prendre en compte. L'inscription gratuite est nécessaire. Un rapport HTML, Word ou PDF est envoyé par courriel. Le rapport HTML propose la poursuite de la recherche sur Google, Bing ou Yahoo. **3** www.doccop.com/web_check.html?nc=42040193

- **Plagium** est un service en ligne de recherche de plagiat sur un texte de 25 000 caractères au maximum pour sa partie accessible gratuitement. Il utilise

les moteurs de Yahoo et de Bing, et offre une présentation des résultats originale **4**.

www.plagium.com

● SeeSources.com propose gratuitement, à titre d'essai, la recherche sur 1 000 mots au maximum.

www.plagscan.com/seesources/analyse.php

● 3YA recherche la phrase soumise dans trois moteurs de recherche : Google Books (livres numérisés), Google Scholar (travaux universitaires) et Google.

<http://michel.bernard.online.fr/plagiat>

● WCopyfind est une application un peu à part à installer sur son poste, qui ne compare que des fichiers à sélectionner de types texte, HTML ou Word (.doc uni-

quement) selon des paramètres choisis. Les fichiers à comparer étant à préciser, la recherche exhaustive sur le Web n'est pas possible. Il peut être utile pour vérifier si des étudiants ont copié les uns sur les autres, ou sur une source que l'on a repérée sur le Net.

Pour le télécharger :

<http://plagiarism.phys.virginia.edu/Wsoftware.html>

Pour les solutions payantes, la facturation se fait soit par l'achat préalable de crédits qui seront débités selon le nombre de mots contenus dans le fichier à analyser, soit par un abonnement de type compte, et quelquefois à l'unité. En voici une liste non exhaustive :

● Turnitin, l'un des plus connus, est utilisé par plus de 10 000 établissements d'enseignement dans 126 pays.

Téléchargement du document

Classement dans des dossiers personnalisables

Trois choix possibles pour le document :

- Exécuter l'analyse et intégrer le document dans la base de données
- Ne pas faire d'analyse, mais intégrer le document dans la base de données
- Exécuter l'analyse, mais ne pas intégrer le document dans la base de données

Analyse terminée
Indication du taux de similarité

Choix d'une vue sommaire ou détaillée

Rapport d'analyse

Accès aux sources trouvées

Rapport d'analyse PDF téléchargé

Mise en évidence des passages repérés en regard de la source probable

Mise en évidence des passages repérés en regard de la source probable

5 La détection par Ephorus

La recherche s'effectue sur internet, mais aussi dans une base de données contenant plus de 135 millions de copies et publications. Turnitin affirme pouvoir détecter les copies achetées, celles issues de sites « anti-sèche » ainsi que les paraphrases entre étudiants. Le service propose différents modules : OriginalityCheck, qui détecte le plagiat uniquement, en précisant le type de sources (site internet, publication, copie d'un autre étudiant...) et le taux de similarité entre la copie et la source trouvée ; PeerMark, qui permet au professeur et à tout étudiant de commenter la copie d'un étudiant, tous les commentaires étant visibles ; GradeMark, qui donne la possibilité à l'enseignant de commenter et d'évaluer la copie selon ses propres critères.

http://turnitin.com/index_i18n.html?lang=fr&pref=1

● **Ephorus**, leader sur le marché européen, bien implanté dans les universités françaises, propose gratuitement un « compte démo ». La base de données de recherche inclut quelques sites d'e-learning 5.

<https://www.ephorus.com/fr>

Citons encore :

● **Compilatio** et **Pomprot**, du même éditeur, l'un nécessitant une licence, l'autre facturé au fichier :

www.compilatio.net/fr

www.pomprot.com

- **Urkund** : www.urbund.com/fr
- **IPlagiarismCheck** : www.iplagiarismcheck.com
- **PlagScan**, la version pro de SeeSources : www.plagscan.com 6
- **EVE2** : www.canexus.com
- **PlagiarismDetection.org** : <http://plagiarismdetection.org>

Conclusion

L'objectif de la lutte contre le plagiat est de favoriser l'honnêteté intellectuelle et de développer chez l'élève cette compétence qui permet de rechercher les ressources nécessaires à un travail demandé, de les analyser, d'y porter un regard critique, d'émettre et de vérifier des hypothèses, puis d'établir les rapports conceptuels et les connexions nécessaires pour étayer son discours – en fin de compte de s'approprier le sujet. À nous de mettre les outils et les méthodes en place et de faire passer le message auprès de nos apprenants. ■

En ligne

L'enquête « De la documentation... au plagiat » menée entre avril et juillet 2007 auprès de 1 100 étudiants et 120 enseignants au sein de trois établissements phares de l'université de Lyon, l'Insa Lyon, l'université catholique de Lyon et l'sara-Lyon :

www.compilatio.net/.../sixdegres-univ-lyon_enquete-plagiat_sept07.pdf

The screenshot displays the PlagScan web interface. On the left, a sidebar contains navigation options like 'Accueil', 'Mon Profil', 'Mes Documents', 'Mes Rapports', 'Mes Commentaires', and 'Mes Sources'. The main area shows a document titled 'Document de travail' with a progress bar indicating 'Analyse en cours'. A list of matches is visible, showing percentages and source information. On the right, a detailed report is shown, with blue callout boxes highlighting specific sections: 'Ajout de document à analyser', 'Lancer l'analyse', 'Rapport d'analyse', 'Rapport d'analyse au format .doc', 'Rapport d'analyse au format PDF', and 'Commentaires et sources possibles'. The report text includes a list of matches and a detailed analysis of the document's content, mentioning the use of Turnitin's database and the identification of plagiarized text.

6 La détection par PlagScan