Plagiat

Le plagiat (du latin plagiarius, celui qui vole l'esclave d'un autre ou vend une personne libre) consiste à s'inspirer d'un modèle que l'on omet délibérément de désigner. Le plagiaire est celui qui s'approprie frauduleusement le style, les idées, ou les faits.

Le phénomène de triche par plagiat est un problème qui s'est rapidement développé au cours de ces dernières années. Au lieu de produire un travail original, certains étudiants préfèrent copier directement le contenu trouvé dans des livres, des encyclopédies, des articles de journaux ou des travaux antérieurs rédigés par d'autres étudiants, voire sur l'un des nombreux sites de triche disponibles aujourd'hui sur Internet – sites offrant du matériel existant parfaitement organisé par sujet et niveau pour un accès plus facile.

Le problème et son ampleur ont été confirmés par une série d'études.

✓ Un rapport datant de 2003 établit par l'agence suédoise Swedish National agency for Higher Education (HSV) affirme que la triche par plagiat représente la majorité des cas traités par les comités disciplinaires des universités et lycées du pays.
L'agence indique également que le nombre de cas a rapidement augmenté ces dernières années et que les cas non enregistrés peuvent se révéler très élevés pour des sujets tels que les thèses.

✓ Dans un rapport pour le congrès de l'association suédoise des enseignants universitaires SULF (Swedish Association of University Teachers), le Professeur Claes Sandgren a déclaré ceci '.. il s'agit d'un problème à la fois juridique et de contrôle. La triche et le plagiat ne sont plus acceptables car ils sont perpétrés par Internet et non par des livres ou des magazines, rendant le contrôle extrêmement difficile et facilitant les « vols » entre étudiants'.

✓ Selon une étude australienne (2000) qui a contrôlé le contenu de 1 751 essais et travaux rédigés par des étudiants de six universités, 1 copie sur 12 (environ 8 %) incluait une quantité significative (25 % au minimum) de texte plagié sur des sources Internet.

Enquête sur le plagiat - septembre 2007

Enquête menée en partenariat avec l'Université de Lyon

Cette étude a été 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 :

- INSA Lyon
- Université Catholique de Lyon
- ISARA Lyon

Cliquez ici pour télécharger les résultats complets de l'étude

Cette enquête révèle notamment que :

- Les étudiants **favorables** à une « **formation à la documentation** » plagient **plus** que les autres.
- **Les formations méthodologiques** menées seules (sans contrôle) ne font que **faiblement baisser** le recours au « copier-coller ».

Le plagiat sur Internet

- **4 étudiants sur 5** (79,7 %) déclarent avoir **recours au "copier-coller"** !
- **9 enseignants sur 10** ont déjà été confrontés au "copier-coller".

Une confusion entre « citer » et « plagier »

- **3 professeurs sur 5** estiment que **les citations sont mal identifiées** dans les travaux des étudiants.
- Plus de **4 élèves sur 5 ont été formés** à la « recherche documentaire ».

Le passage au numérique

- **4 travaux rendus sur 5** existent sous forme **numérique**, mais **3 travaux sur 5 sont rendus** sur papier.
- Les E.N.T. (Environnements Numériques de Travail) sont utilisés de manière privilégiée pour collecter les travaux par seulement 6% des enseignants.

Egalement dans cette étude

- Des analyses croisées permettant d'établir des typologies d'étudiants en fonction de leurs comportements :
    - Fréquence du recours au plagiat
    - Quantité de plagiat dans un devoir type
- Un comparatif des résultats de l'enquête menée en 2006 avec les résultats d'aujourd'hui

Cette étude similaire menée en 2006 auprès de 1200 étudiants et enseignants issus de l'enseignement supérieur, et plus particulièrement de grandes écoles donnent des résultats similaires.

Cliquez ici pour télécharger les résultats complets **2006**

Il ressort notamment à la lecture de cette étude que :

Internet, principale source de documentation des étudiants

- **Tous les étudiants utilisent Internet pour se documenter** quand seulement 1 étudiant sur 2 se rend encore en bibliothèque.

Une tendance forte chez les étudiants : le "copier-coller" sur Internet

- **3 étudiants sur 4** (77 %) déclarent avoir **recours au "copier-coller".**

Quantification du “copier-coller”

- **3 travaux sur 4** (73,7%) contiennent au moins un passage copié à l’identique sur Internet.
- **7 étudiants sur 10** (69,8 %) pensent qu’**un devoir type contient au moins 1/4 de textes recopiés sur Internet**.
- Parmi ceux qui incluent au moins 25% de “copier-coller" dans leurs devoirs, **3 étudiants sur 5** (61,3%) **considèrent avoir « rarement » recours au “copier-coller”**.

Avec le développement d'Internet et des nouvelles technologies, le phénomène du plagiat s’est beaucoup développé, en particulier dans le milieu universitaire. De nombreux enseignants cherchent alors des moyens efficaces pour lutter contre le plagiat. Il existe plusieurs solutions.

Les moteurs de recherche

Tant que le plagiat restait un phénomène, pour détecter un éventuel plagiat, la première solution consistait à chercher sur un moteur de recherche des mots ou des phrases clés du texte en question, afin de voir si l'on retrouve un texte potentiellement plagié.

Cette méthode peut s'avérer particulièrement efficace dans le cas d'un élève ayant totalement recopié un article trouvé sur Internet. Par contre, si l'élève a plagié de multiples sources, ou seulement une partie de son devoir, une telle recherche peut vite devenir fastidieuse.

Mais la taille exponentielle d'Internet rend cette méthode futile, d'autant plus que les moteurs de recherche n'indexent qu'une partie des sites existants. En conséquence, le professeur doit faire appel à plusieurs moteurs, ce qui augmente proportionnellement le temps passé. L’utilisation des moteurs de recherche courants tels que Altavista, Google et Yahoo, il y a de grandes chances pour qu'il n'obtienne aucun résultat, car ce type de contenu est généralement masqué dans des fichiers compressés (.zip, .arc, .lha, etc.) qui doivent être décompressés avant que toute comparaison soit rendue possible. En outre, ce matériel compressé n'est souvent accessible que sur des sites payants ou protégés par mot de passe, compliquant encore plus la recherche de la source originale et la rendant parfois coûteuse, du fait des centaines de sites payants proposant des travaux scolaires. Enfin les sites qui affichent des pages dynamiques, ou qui sont derrière une identification (un login) ne peuvent pas être testés. Certains sites spécialisés dans la vente de travaux d'élèves ne vous les montrent pas avant que vous les payiez !
Même si certains moteurs se révèlent efficaces, le temps passé et les efforts nécessaires pour contrôler les sources des étudiants sont inconcevables étant donné la multiplication des travaux à contrôler.
Pour combler le tout, Internet est loin d'être l'unique source disponible pour le plagiaire potentiel, ce qui complique d'autant plus le problème. En effet, toute sorte de matériel publié (presse scientifique, journaux, livres, encyclopédies, etc.) ainsi que les travaux rédigés par d'autres étudiants (famille ou amis) sont des ressources exploitables par quiconque souhaite plagier.

Les logiciels de détection du plagiat

Pour lutter contre ce phénomène grandissant, plusieurs logiciels sont apparus ces derniers temps. Ils ont des caractéristiques sensiblement identiques, mais leur efficacité varie de façon considérable.

Ces logiciels vont de la simple comparaison de deux documents, à la rechercher automatique de sources issues de l'Internet pour les plus performants. Suivant leur degré d'aboutissement, ils sont en mesure de traiter un nombre plus ou moins important de formats de fichiers, les plus répandus étant les fichiers Word, PDF et html.

Pour permettre une évaluation des logiciels anti plagiat, il semble nécessaire de faire plusieurs distinctions :

Type de fonctionnement :

- ceux qui travaillent sur un serveur distant.
- ceux qui peuvent être installés directement sur la machine de l'utilisateur, et utilisés en local.

Le premier type de logiciel semble a priori le plus efficace car il dispose bien souvent d'une base de données de référence gigantesque, qui s'enrichit à chaque fois qu'un nouveau document lui est soumis pour analyse, le serveur l'incorporant alors à ses documents de référence. A noter aussi l'intérêt d'avoir la possibilité pour le professeur de demander à ses étudiants de rendre les devoirs sur une adresse mail spécifique : cela facilite grandement le travail du professeur qui, une fois le délai dépassé, reçoit un compte rendu global.

Langue de fonctionnement :

- Français
- Etranger

A vrai dire aucun des systèmes de langue anglaise n'a donné satisfaction.

Coût du système :

- Gratuits
- Payants

Les logiciels gratuits, lorsqu'il a été possible de les faire fonctionner (!) ont systématiquement donné des résultats médiocres.

Pour mener à bien ce test, j'ai repris une copie d'étudiant, portant sur le sujet 'Conséquences de l'élargissement de l'Europe, au niveau des I.D.E'. J'ai inclut dans cette copie 6 passages copiés sur Internet dans des sources diverses (ces passages sont en rouges dans le document de test, avec une note en bas de page indiquant la source) et 1 passage copié puis modifié (des pluriels ont été changés en singulier). Ce passage est en bleu avec également une note de bas de page, indiquant la source. Au total 338 mots plagiés sur 2476 mots soit 13.65 %.
Document : Travail.pdf joint.

Les sources utilisées sont les suivantes :

http://www.educnet.education.fr/insee/comext/effets/effets21.htm
http://enisse.kharroubi.free.fr/index2.html
http://www.notre-europe.eu/fileadmin/IMG/pdf/Etud24-fr.pdf
http://www.lepotentiel.com/afficher_article.php?id_edition=&id_article=51958
http://ec.europa.eu/archives/commission_1999_2004/verheugen/speeches/sp14012002fr.htm
http://www.journaldunet.com/management/dossiers/040433europe/
http://findev.blogspot.com/

Schéma de fonctionnement

Le logiciel commence par traiter le document à analyser afin de ne conserver que son contenu textuel.
Plusieurs modes de fonctionnement sont alors possibles :

- Recherche de mots clefs dans le texte.
- Sélection d'un certains nombre de mots au hasard, qui seront ensuite recherchés.
- Sélection aléatoire de phases complètes (suite de mots).

La méthode la plus logique consiste à sélectionner au hasard des passages du texte à analyser et soumettre ces passages aux différentes sources disponibles. Le texte ne sera pas intégralement analysé, ce qui permet de gagner du temps. Si le texte sélectionné est retrouvé dans une des sources, alors tout le texte sera vérifié par rapport à cette source.
Tous les extraits retrouvés à l'identique sont considérés comme plagiés. Si un extrait est en fait une citation ou une phrase très courante, cet extrait peut être ignoré. Il ne sera plus pris en compte dans le calcul du coefficient de plagiat.
En fonction des algorithmes de recherche, deux problèmes sont fréquemment rencontrés :

- La recherche se fait sur des phrases ou partie de phrases similaires : Il est évident que si l'étudiant modifie la structure, la forme, une partie des mots, utilisant des synonymes par exemple, le plagiat ne sera plus détectable : mais où commence et où finit le plagiat ?
- Il n'est pas rare qu'une suite de mots posés les uns après les autres, de façon aléatoire, se retrouvent dans divers document sur Internet, sans pour autant qu'on puisse parler de plagiat. Ce sera souvent vrai pour le titre d'un exposé ou pour des phrases très générales su type 'Les conséquences économiques de la hausse de l'euro'. La classification en plagiat ou non dépendra de la puissance de l'algorithme.

Les sources de recherches sont de trois types :
Internet. Le logiciel fonctionne comme un méta moteur : il interroge différents outils de recherches et centralise les résultats obtenus. Cette recherche, en fonction des logiciels peut se faire sur l'Internet public (ouvert à tous) ou l'Internet clos (accessible sur abonnement).
Contenu publié : Livres et revues accessibles sous format numérique.
Base de documents ou de données du logiciel : Archives internes regroupant des travaux d'étudiants, thèses, mémoires dissertations, exposés etc.
Plus la base de recherche sera importante, meilleure sera la qualité des résultats.

Logiciels testés :

Pour mener à bien cette étude, qui ne se veut pas exhaustive, j'ai sélectionné tous les logiciels existants sur le marché.
Une étude publiée en Septembre 2007 en Angleterre (doc PDReview-Reportv1_5.pdf) donne certains résultats sur lesquels il est déjà possible de se reposer :

Category	Docoloc	Ephorus	My Drop Box / SafeAssign	Turnitin	Urkund
Installation / Configuration	6.3	4.5	4.4	7.7	6.9
Stability of product	2.1	2.4	1.7	2.8	2.8
Databases searched	1.5	0.9	4.9	6.8	3.9
Usability	4.4	3.1	5.7	5.9	5.0
Upload options	3.3	3.1	4.1	5.2	4.6
Output options	3.3	1.4	5.9	6.1	3.5
Support	1.1	1.4	2.1	3.5	2.6
Company background	0.2	0.4	1.2	3.2	1.0
Company associations	0.0	0.0	1.5	2.5	2.5
Pricing	0.0	0.0	0.0	0.0	0.0
Accessibility	0.0	0.0	0.6	2.5	0.0
Authentication	1.0	1.8	1.8	2.5	1.8
Legal considerations	0.0	1.3	1.2	2.5	2.5
Privacy	1.8	2.3	2.7	3.3	3.0
Total	**24.9**	**22.5**	**37.6**	**54.5**	**39.9**

Un mail envoyé à Docoloc et à MyDropBox afin d'obtenir des codes de test, étant resté sans réponse, il ne feront pas partie de cette étude. Les trois autres : Ephorus, Turnitin et Urkund seront testés.
A coté de ceux-ci, une dizaine d'autres logiciels dont voici la liste :

Logiciels gratuits

- CopyTracker logiciel libre de détection du plagiat (peut comparer des rapports)
- Plagium, service en ligne de recherche de plagiat sur un texte court
- Noplagia, logiciel libre (GPL) de détection du plagiat écrit en perl. (les sources)
- http://www.lucboyer.com/renifleur/

Logiciels commerciaux

- Compilatio.net
- http://www.urkund.com/fr/
- http://www.turnitin.com/static/home.html
- http://www.canexus.com/eve/
- http://www.iplagiarismcheck.com/static/pricing.html
- http://www.ephorus.nl/

http://copytracker.org/ logiciel libre de détection du plagiat (peut comparer des rapports)

~~Payant~~ Gratuit. Langue : français. Installé en interne

Après installation et lancement

plantage.

Warning: fsockopen() [function.fsockopen]: unable to connect to http://www.google.com/search?q=pays+investissements+effets+dans+investisseurs+:80 (Unable to find the socket transport "http" - did you forget to enable it when you configured PHP?) in **C:\Program Files\Copytracker\htdocs\googleCrawler.class.php** on line **39**

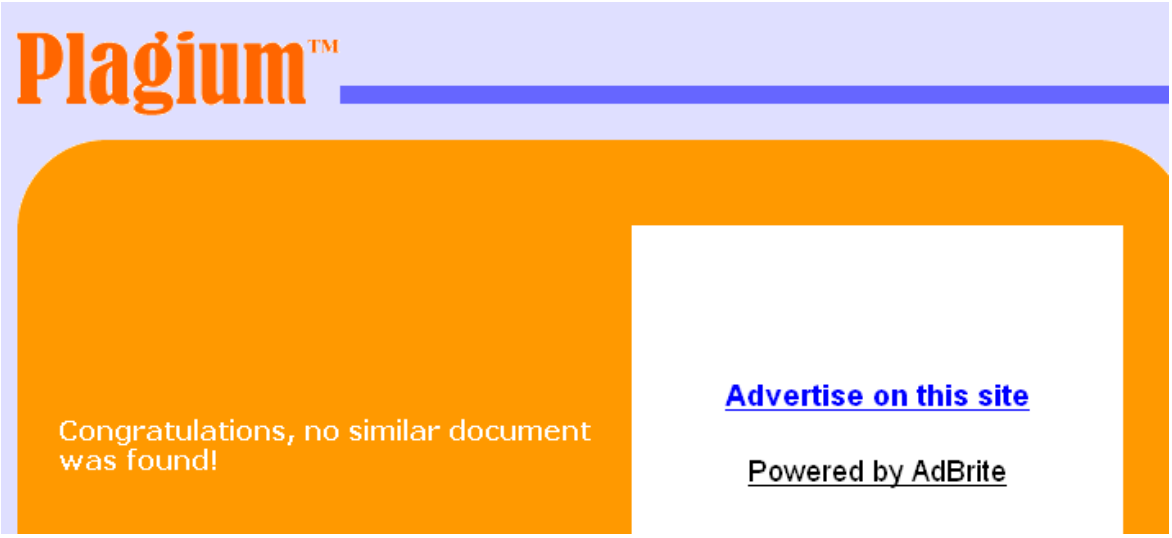
Mail envoyé.
Attente réponse.

http://www.plagium.com/

~~Payant~~ Gratuit. Langue : anglais. Installé sur serveur externe

On entre le texte dans la fenêtre (texte limité à 1000 caractères).

Résultat : C'est parfait (il y a la moitié du document qui provient d'Internet !)



No Plagia

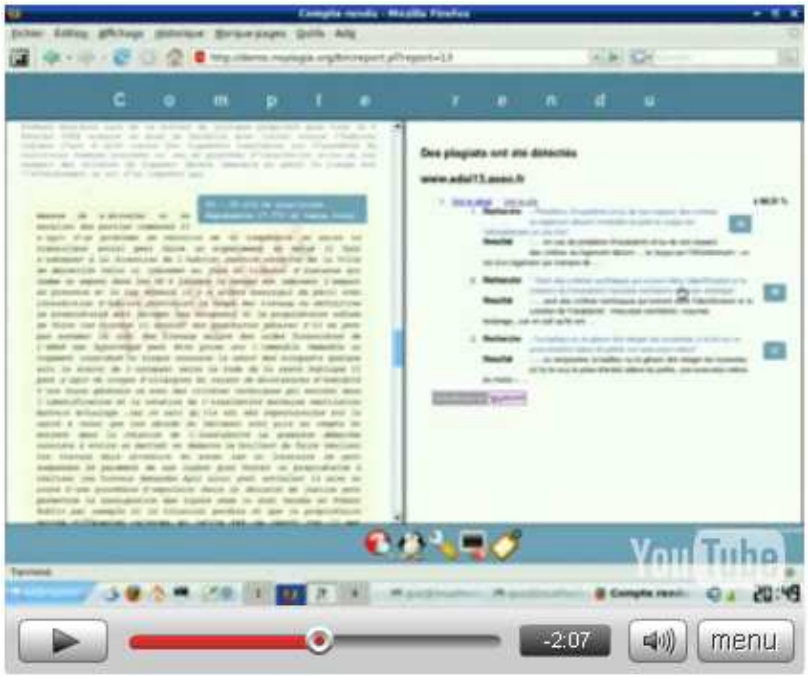
http://noplagia.org/ logiciel libre (GPL) de détection du plagiat écrit en perl. (les sources)

~~Payant~~ Gratuit. Langue : français. Installé sur serveur externe

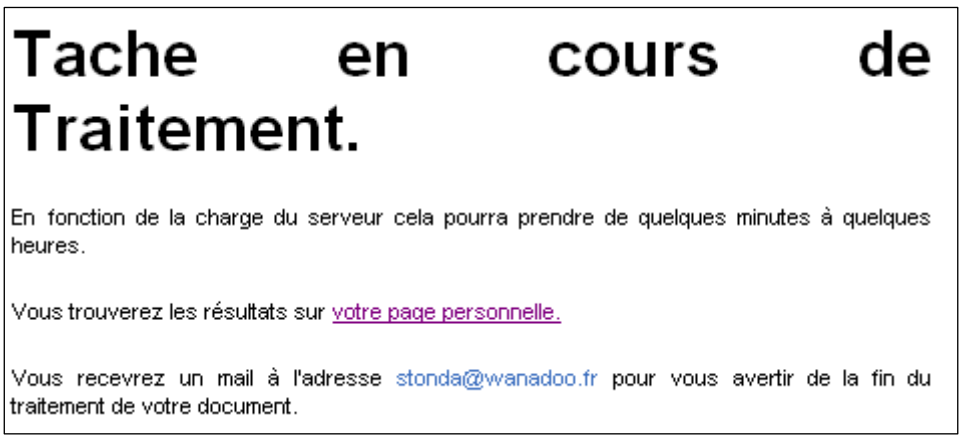
Test sur http://demo.noplagia.org

login : duguest et mot de passe : noplagiatest

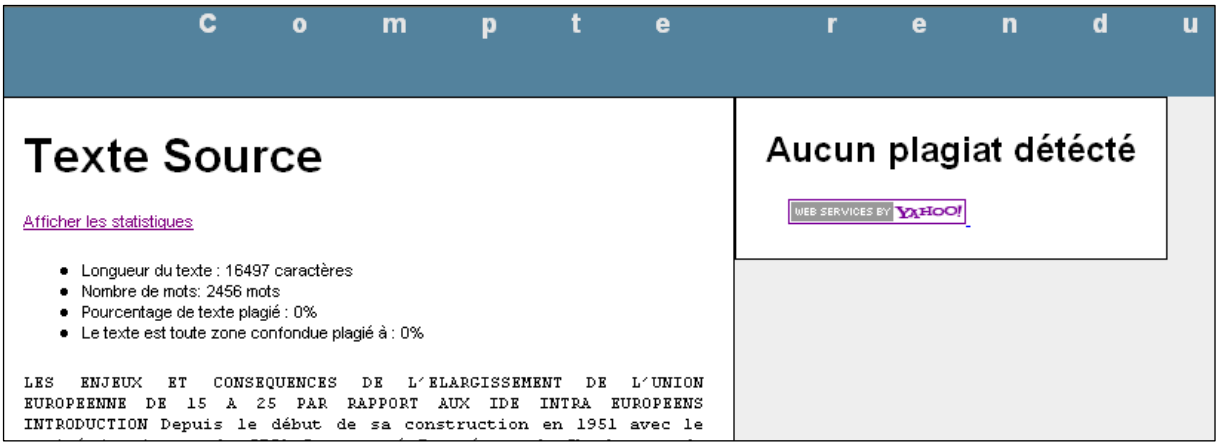
Démo visible sur YouTube : http://www.youtube.com/watch?v=GRP0ELVOe_4



On entre le texte en brut dans la zone créé à cet effet. Le traitement se lance



Après quelques minutes, un mail nous informe que le test est terminé et le résultat disponible :



C o m p t e r e n d u

Texte Source

Afficher les statistiques

- Longueur du texte : 16497 caractères
- Nombre de mots: 2456 mots
- Pourcentage de texte plagié : 0%
- Le texte est toute zone confondue plagié à : 0%

LES ENJEUX ET CONSEQUENCES DE L'ELARGISSEMENT DE L'UNION EUROPEENNE DE 15 A 25 PAR RAPPORT AUX IDE INTRA EUROPEENS INTRODUCTION Depuis le début de sa construction en 1951 avec le

Aucun plagiat détécté

WEB SERVICES BY YAHOO!

Résultat qui ne semble effectivement pas satisfaisant !

Le renifleur

~~Payant~~ Gratuit. Langue : français. Installé sur serveur externe

http://lucboyer.com/reniflard

Le renifleur de plagiat

En cours de modification.

Bonjour,
Le moteur de recherche utilisé pour la recherche a modifié son CSS, ce qui rendait la version présente inopérante. Bien que j'aie "réparé" cette altération, j'ai pris la décision de reconstruire le programme dans son intégralité, d'en faire un projet éducationnel, évolutif... et gratuit.
La nouvelle version ne sera pas cependant disponible avant janvier 2008, en version bêta.
J'ai donc pris la décision de retirer la présente version afin de ne pas "nuire" à sa réputation.
Désolé pour le délai,
Luc Boyer

Dans l'ancienne version :

Le renifleur de plagiat supporte les documents de Microsoft Word (doc), Rich Text (rtf) ainsi que les fichiers texte (txt). Évidemment, une connexion Internet est obligatoire et un lien haute vitesse est recommandé bien que non impératif. Avec la version complète, vous pouvez vérifier autant de documents que vous le désirez, en autant que ceux-ci se retrouvent dans un même dossier.

Le renifleur de plagiat est un méta moteur et utilise des moteurs de recherche sur Internet. Plusieurs chaînes de mots sont extraites du document à traiter et sont ainsi recherchées afin de trouver une source sur Internet. Si le résultat est positif, le renifleur de plagiat affichera les hyperliens trouvés pour chaque chaîne de mots.

Vous pouvez paramétrer le processus de recherche en ajustant la taille de la chaîne de mots ainsi que le nombre de tests requis par document.

A priori, dans l'ancienne version en tous cas, la détection se fait sur la fréquence des occurrences identiques.

http://www.canexus.com/eve/

Payant ~~Gratuit~~. Langue : anglais. Installé en interne

Votre nom d'usager est : xxxxxxx
Votre numéro ou code d'usager est : XXXXXXXXXXXXXXX

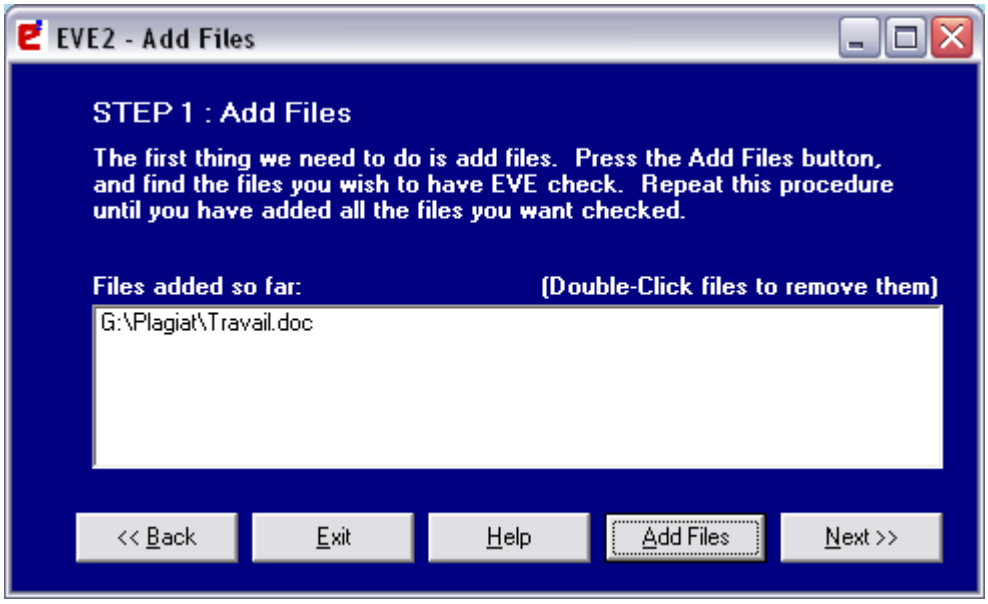
Purchase EVE2

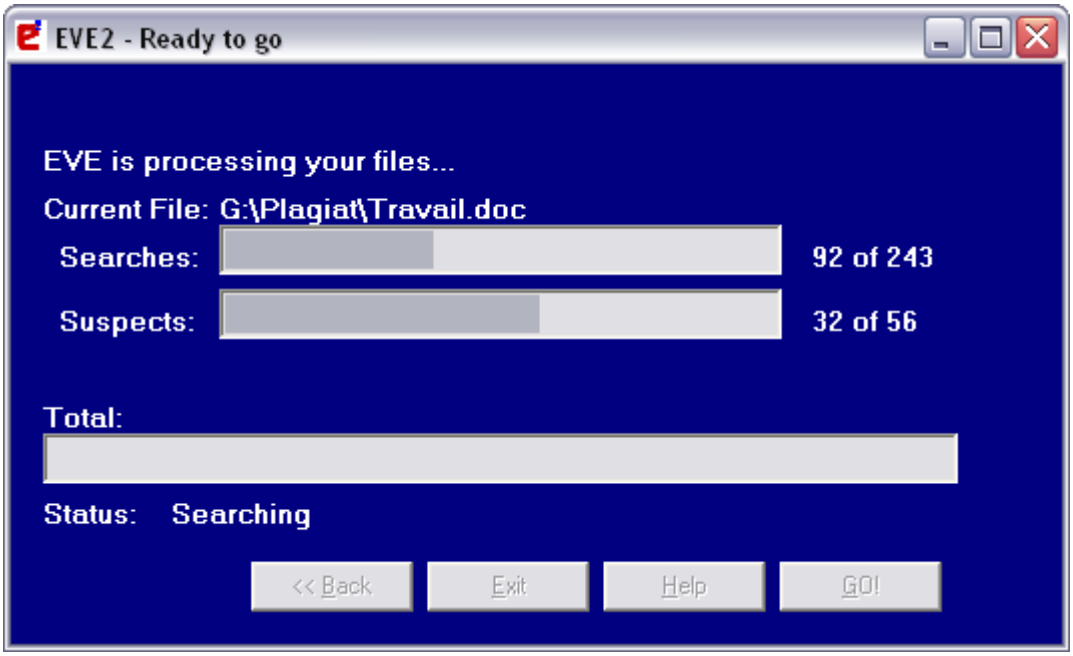
EVE can be purchased online for only $29.99 per licence. **This is a one time cost to purchase EVE2.** If EVE2 is updated, you will be eligible for a free update. If a new version (ie:EVE3) is released within 60 days of your purchase, you will receive a free upgrade. If a new version is released later than 60 days from your purchase, you will be eligible for a special upgrade price.

If you purchase online, you will receive your User-name and Registration Code immediately via an automatically generated email so that you can start using EVE2 right now!

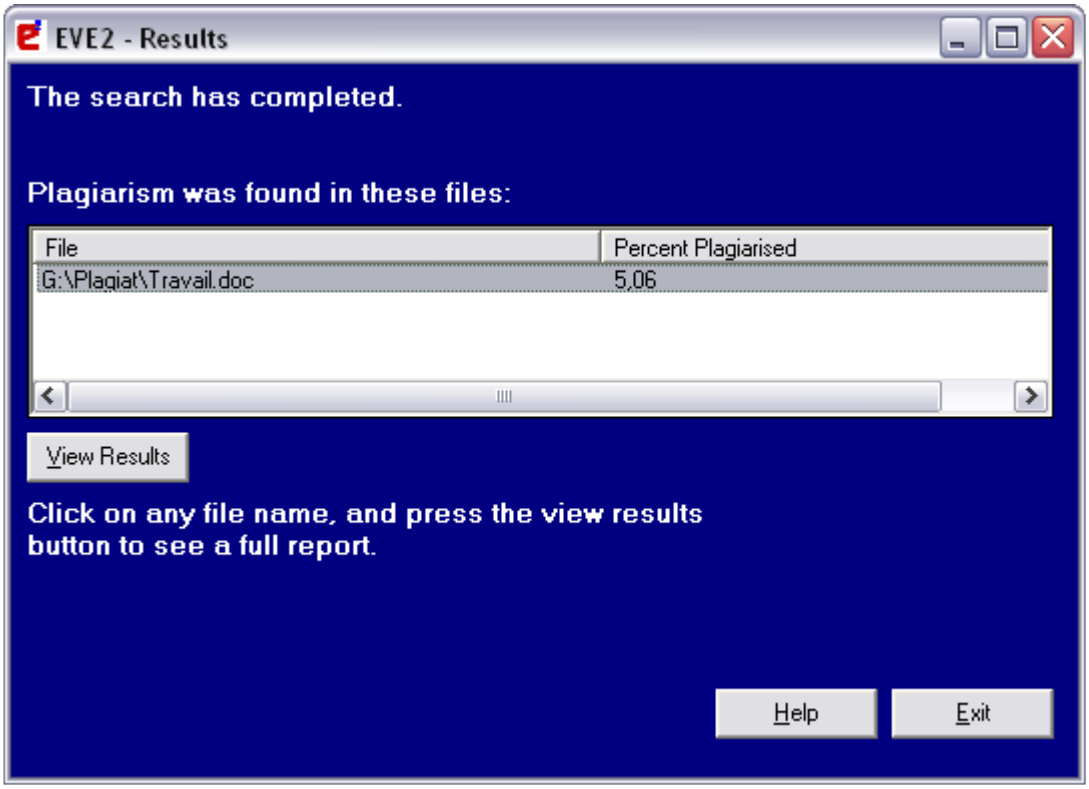
EVE comes with a 10-Day money back guarantee! If you are not satisfied with EVE, you simply need to fill out a very small form on our website, and we will refund your full purchase price!

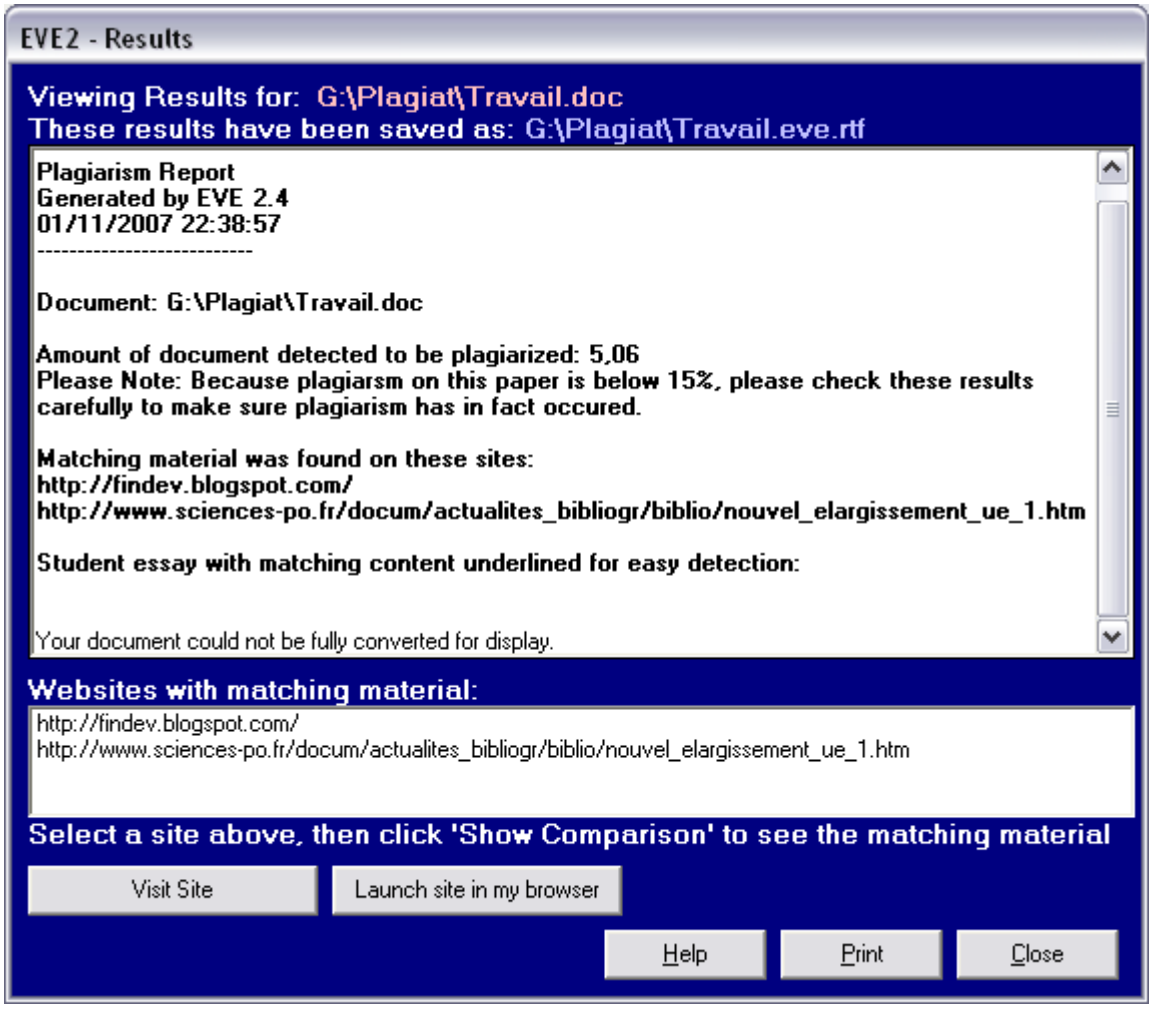
Licence Information: Each professor or teacher must purchase their own copy of EVE. A copy of EVE may be used by one teacher for all of the essays submitted in all of the classes they teach. It may not be used to check essays for another teacher unless that teacher purchases a licence as well.





Les résultats s'affichent :





Difficile à interpréter.

Peu de correspondance avec la réalité :

http://findev.blogspot.com/
http://www.sciences-po.fr/docum/actualites_bibliogr/biblio/nouvel_elargissement_ue_1.htm

comparé aux vraies sources :

http://www.educnet.education.fr/insee/comext/effets/effets21.htm
http://enisse.kharroubi.free.fr/index2.html
http://www.notre-europe.eu/fileadmin/IMG/pdf/Etud24-fr.pdf
http://www.lepotentiel.com/afficher_article.php?id_edition=&id_article=51958
http://ec.europa.eu/archives/commission_1999_2004/verheugen/speeches/sp14012002fr.htm
http://www.journaldunet.com/management/dossiers/040433europe/
http://findev.blogspot.com/

Par ailleurs pas de possibilité de voir le document avec les zones plagiées :

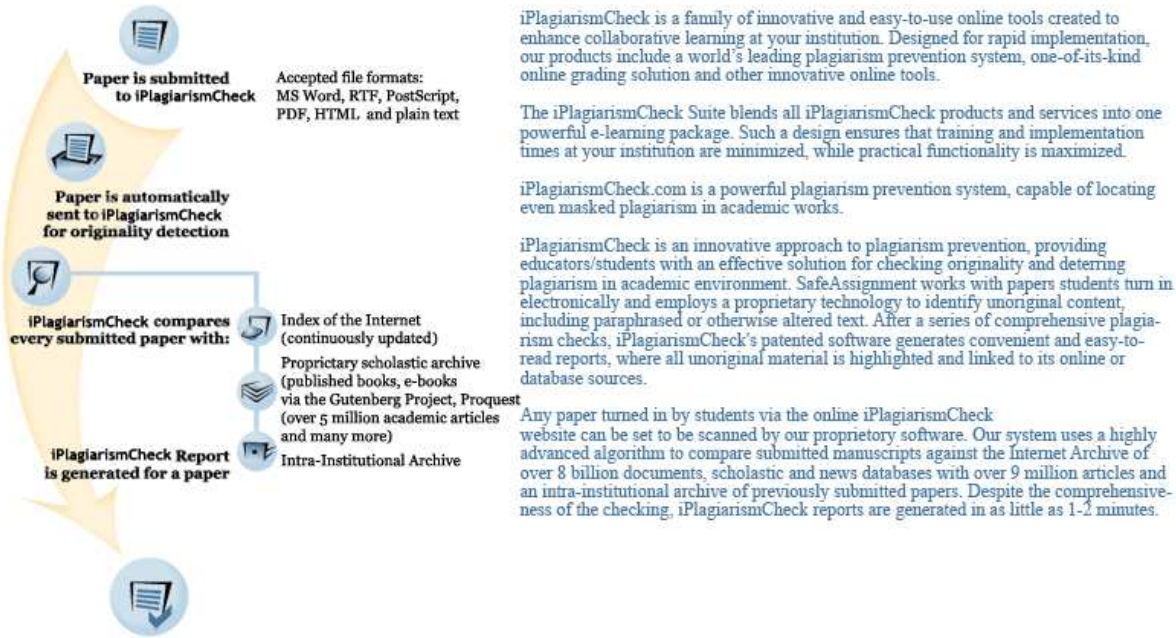
“Your document could not be fully converted for display.”

http://www.iplagiarismcheck.com/static/index.html

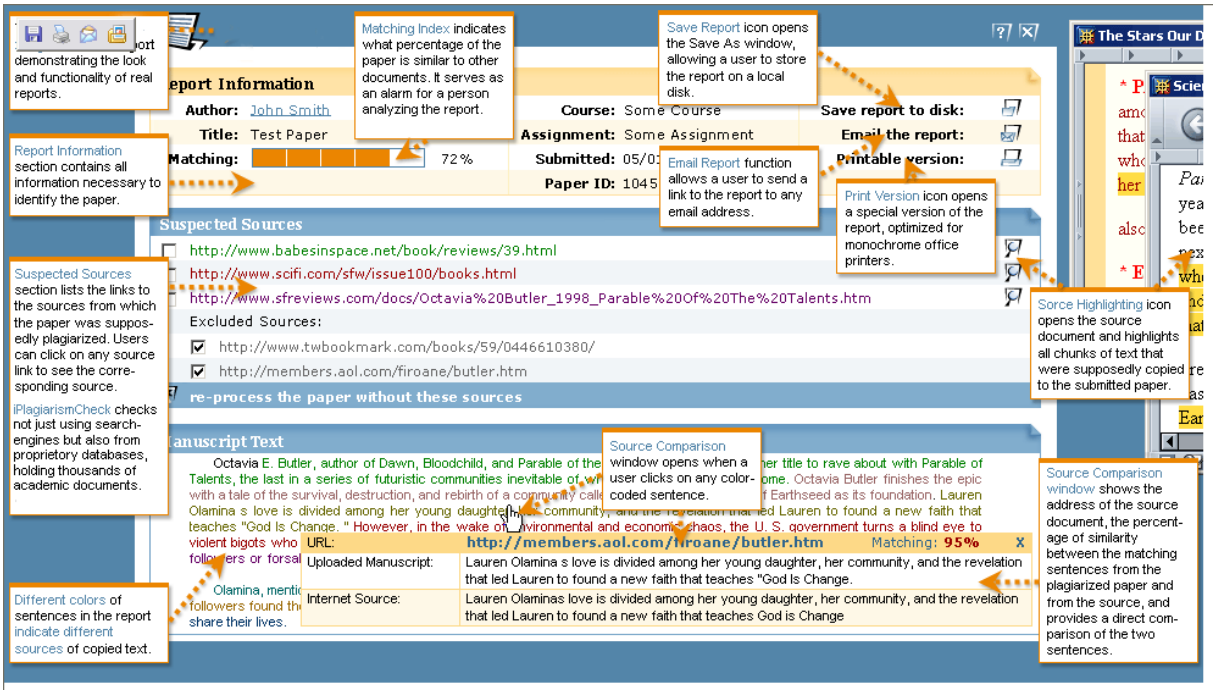
Payant ~~Gratuit~~. Langue : anglais. Installé sur serveur externe

En fonction du contrat souscrit, il est envoyé un mail avec les codes d'accès :
Thank you for signing up for our **Basic Edition** package. Your current subscription entitles you to **Single document submission** **(on an 'Urgent Priority' basis)**. Your 2CO Order Number is: **3830460948**
Il suffit alors de soumettre le document en utilisant ce code.
Site actuellement en modification concernant les entrées et résultats.

Fonctionnement :



Le résultat théorique :



Intérêt : chaque zone plagiée devient interactive et point vers la source.
Problème : le pointeur va généralement sur la page d'accueil du site, d'où difficulté à retrouver l'origine exacte.
Pour l'instant il s'agit d'un fichier html, moins lisible que ce qui est prévu sur le site :

http://www.checkyourassignments.org/3830460948/Travail_54313383.html

iPlagiarismCheck Originality Report

Travail.doc by Duguest D. Dd

Processed on 10-30-07 2:08 PM CDT ID: 54313383 (Word Count:) 2500

Overall Similarity Index: 22%

9% match (internet)
http://www.notre-europe.asso.fr/IMG/pdf/Etud24-fr.pdf

1% match (internet from 11/22/05)
(11-22-05) http://enisse.kharroubi.free.fr/index2.html

1% match (internet)
http://www.industrie.gouv.fr/observat/bilans/pdf/Chap2_01.pdf

Plusieurs liens indiqués sont inactifs, erronés, ou n'existent plus.

Les investissements directs etrangers sont un moyen d'acceder a des secteurs tres innovants des pays developpes, comme les biotechnologies, les nanotechnologies, la pharmacie... Mais l'acces au marche reste l'element majeur. La zone du monde qui a connu la plus forte croissance du PIB en milliards de dollars entre 2002 et 2005, ce n'est pas la Chine, ni les Etats-Unis, mais l'Europe de l'Ouest. Il n'y a pas beaucoup de raisons que le niveau des flux d'investissements directs s'inflechisse. La croissance mondiale reste bien orientee, les financements sont abondants, les entreprises continuent de s'internationaliser. [9]

La dernière phrase du texte est considérée comme non plagiée !

LES ENJEUX ET CONSEQUENCES DE L'ELARGISSEMENT DE L'UNION EUROPEENNE DE 15 A 25 PAR RAPPORT AUX IDE INTRA-EUROPEENS :

INTRODUCTION :

Depuis le debut de sa construction en 1951 avec le traite instituant la CECA (Communaute Europeenne du Charbon et de l'Acier), l'Union Europeenne (autrefois appelee Communaute Economique Europeenne) n'a cesse de s'agrandir, pour atteindre aujourd'hui le nombre de 27 pays membres (en 2007). Cet elargissement europeen, accompagne de quelques institutions (BCE...), a facilite la circulation d'IDE (investissement directs a l'etranger) intra europeens.
Les investissements directs a l'etranger consistent a injecter des fonds etrangers dans l'economie d'un pays dans le but de, soit y creer des entreprises nouvelles, soit de racheter une entreprise deja existant, tout ce processus devant aboutir a l'acquisition du pouvoir de decision effectif dans la gestion de l'entreprise. Par convention, une relation d'investissement direct est etablie des lors qu'un investisseur acquiert au moins 10 % du capital social de l'entreprise investie.[1]
L'elargissement de l'Union Europeenne aura des consequences sur les IDE intra-europeens, des pays deja membres vers les pays de l'Europe Centrale et Orientale ainsi que sur les economies des pays, qu'ils soient emetteurs ou recepteur.
Dans un premier temps nous verrons les differents facteurs qui influencent les flux d'IDE puis nous etudierons les effets du passage du statut de pays candidats a celuis de pays membres Ensuite, nous analyserons la hausse des flux d'IDE intra-europeens et nous verrons les effets negatifs et positifs de ces IDE sur l'economie des pays..

Sur cette vérification, beaucoup de problèmes :

Note 1 détectée mais source erronée.
Note 2 détectée mais source erronée.
Note 3 OK
Note 4 OK
Note 5 OK
Note 6 non détectée
Note 7 Phrase modifiée : pluriels mis au singulier. La partie de la phrase avec les fautes n’est pas retrouvée.
Note 8 Détectée mais avec un lien pointant vers le document source.
Note 9 non détectée

URKUND

Payant ~~Gratuit~~. Langue : français. Installé sur serveur externe. Professeur envoie les copies ou demande à ses étudiants de la faire.

Contact : Jérémy Alliaume. Tel +32 (0)494 651 812 jeremie.alliaume@urkund.com

Création d'un code et mot passe.
Nom d'utilisateur : dduguest
Mot de passe : kwedwl28

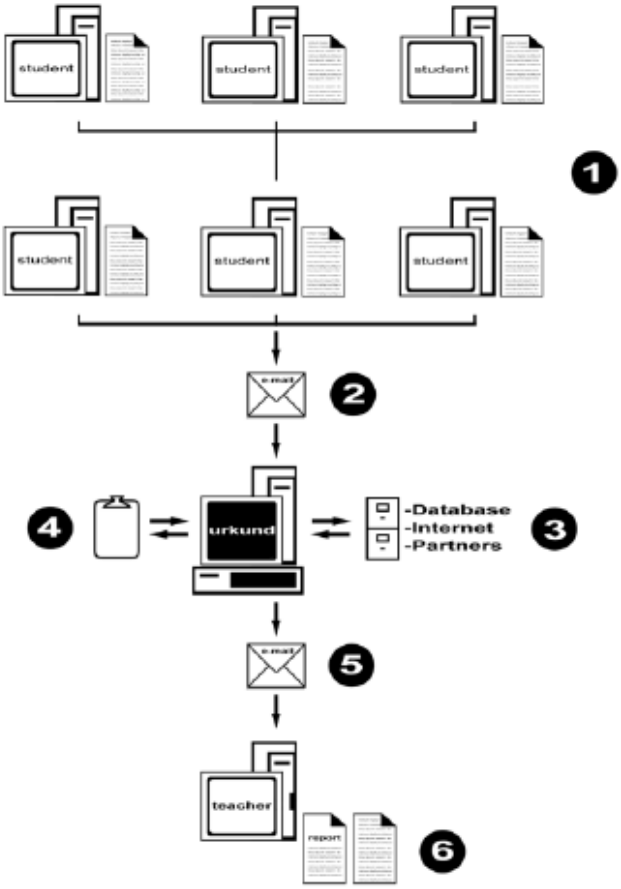
On envoie (le professeur ou l'étudiant) le document à une adresse mail avec une deadline pour la remise des documents. Cette limite atteinte, le système vérifie et envoie le résultat au professeur.

(1) L'étudiant rédige son travail sur PC ou Mac, dans un traitement de texte.

(2) L'étudiant rend sa copie en l'envoyant à l'adresse e-mail que son professeur a reçue d'Urkund.

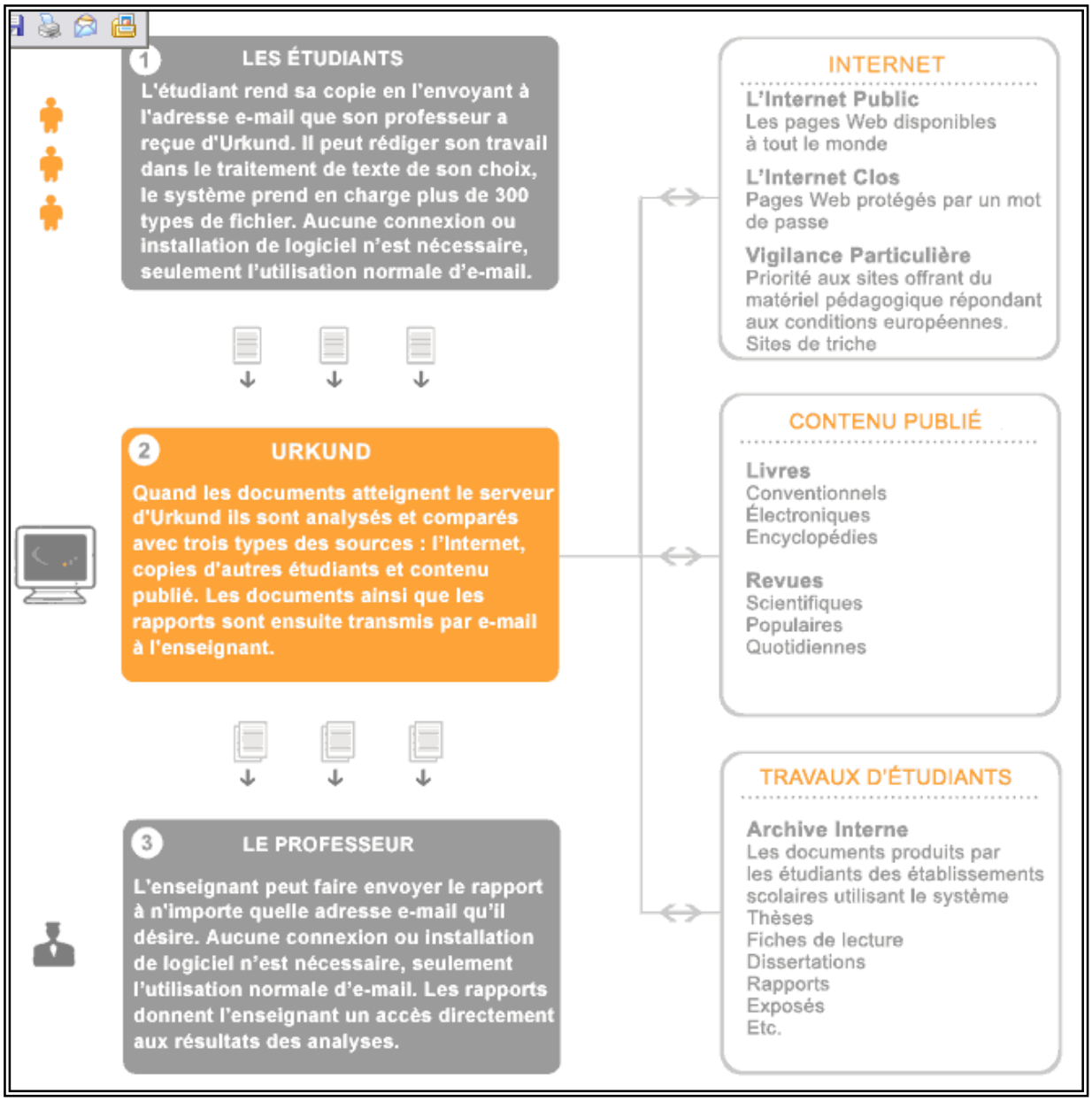
(3) Avant de parvenir à l'enseignant, la copie transite par Urkund, où elle est ajoutée à la base de données, avec le consentement de l'étudiant. Elle est ensuite comparée avec le contenu de la base (sections d'encyclopédies, articles de magazines ou de journaux, livres et sources Internet soigneusement sélectionnées). Le système recherche si la copie est identique en tout ou partie aux sources comparées. En cas de similitudes, il génère un rapport mentionnant la section faisant l'objet de plagiat et le document d'origine.

(4) Urkund consulte les paramètres personnels de l'enseignant afin de trouver son adresse e-mail.

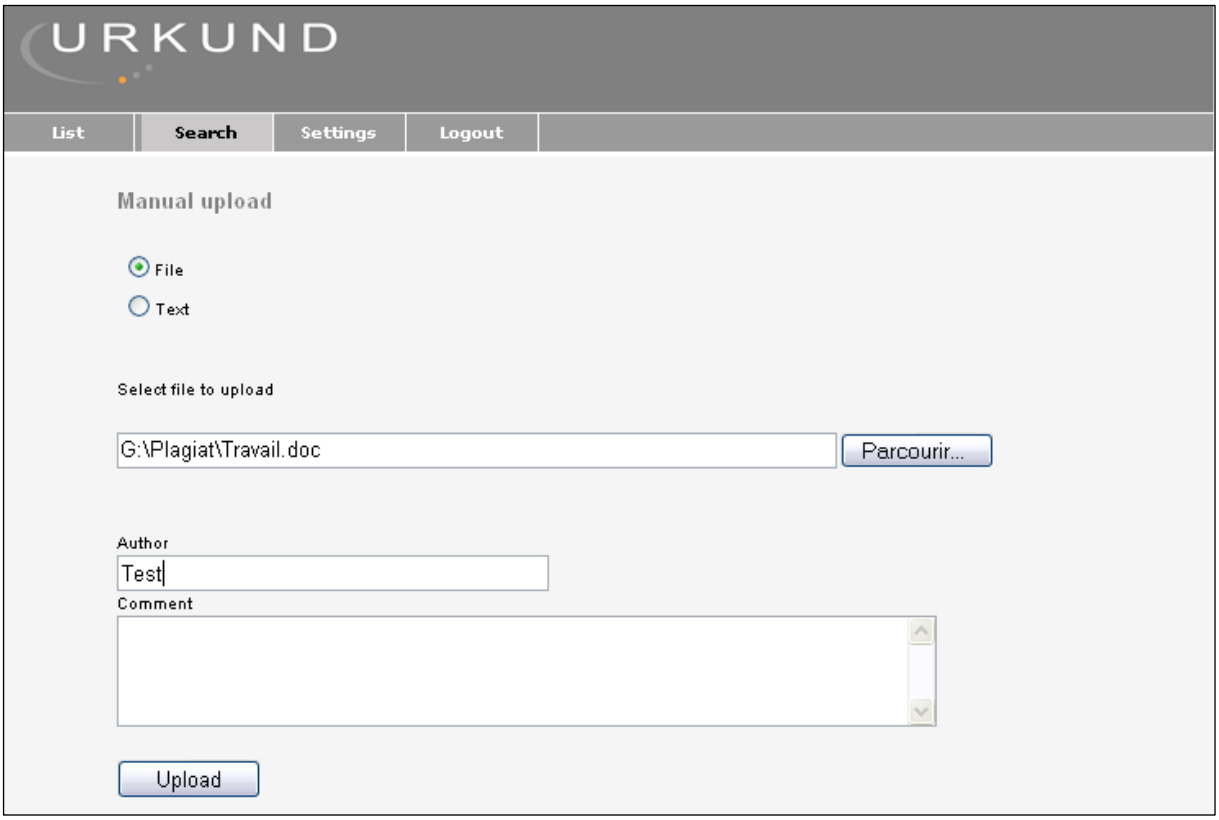


(5) La copie de l'étudiant et un rapport sont ensuite transmis par e-mail à l'enseignant.

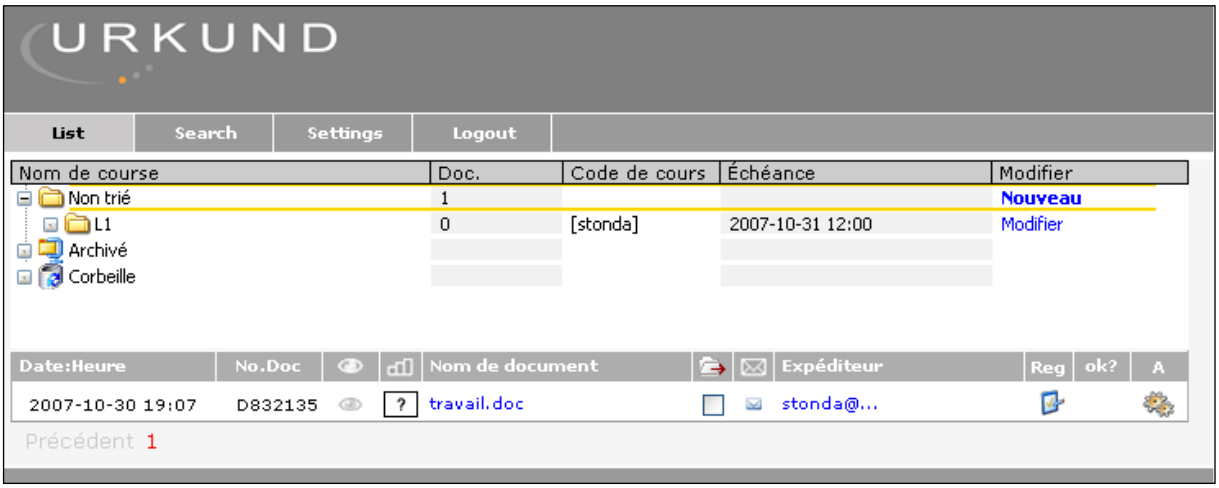
(6) L'enseignant reçoit la copie et le rapport. A partir de ce dernier, il peut facilement déterminer s'il s'agit ou non d'un cas de plagiat.



Import d'un document pour vérification.



Il apparaît dans la liste, en attente de traitement.



Résultat très surprenant pour ce logiciel considéré comme le meilleur sur le marché avec Turnitin. Le taux de plagiat est considéré comme étant très faible.

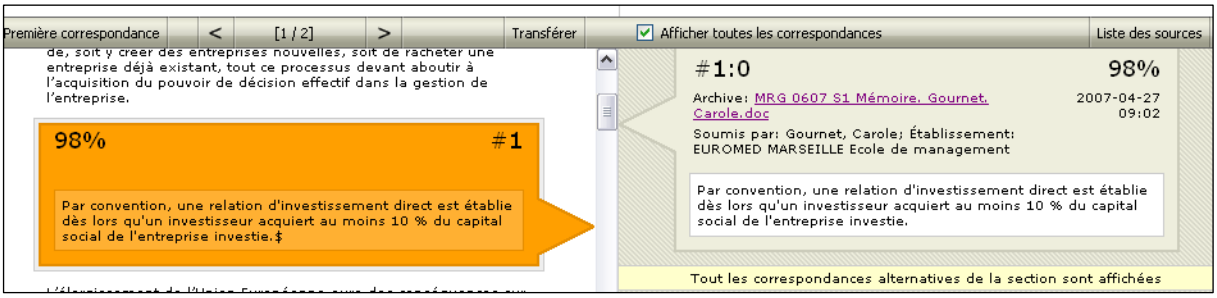
Les résultats se trouvent à cette adresse : http://a2.urkund.se/view/726120-826712-384601
Liste des sources retrouvées par Urkund :

Liste des sources

Type de source	URL/Nom de fichier	Mémorisé en cache
Source externe/Internet	http://www.delta.ens.fr/abstracts/200426.html	2007-10-30 07:28
Source externe/Internet	http://enisse.kharroubi.free.fr/index2.html	2007-10-30 07:28
Source externe/Internet	http://www.insee.fr/fr/nom_def_met/definitions/htm...	2007-10-30 07:28
Source externe/Internet	http://www.industrie.gouv.fr/observat/bilans/bord/...	2007-10-30 07:28
Source externe/Internet	http://www.educnet.education.fr/insee/comext/effet...	2007-10-30 07:28
Archive	MRG 0607 S1 Mémoire. Gournet. Carole.doc	2007-04-27 09:02

Ici les sources 1 et 2 sont identiques. Ce qui est surprenant est que bien que les sources soient identifiées, elles ne pointent pas vers notre document.
Les sources 3 et 6 font référence à la même phrase du document.
Les sources 4 et 5 n'ont pas de vraies correspondances dans notre document.

Contrairement à certains logiciels, Urkund n'affiche pas toutes les correspondances, afin d'éviter, comme il a été dit plus haut, que des phrases banales soient considérées comme plagiées.
Lorsqu'il détecte une similitude elle est affichée de la façon suivante :



La société Urkund s'est saisie du problème et doit me donner une réponse rapide

----- Original Message -----
From: Jeremie Alliaume
To: M. Duguest
Sent: Monday, November 05, 2007 6:12 PM
Subject: Re: Re: Fw: [urkund] 2% similarity - stonda@wanadoo.fr

Mr. DUGUEST ,

Vous avez certainement raison concernant la multitude des sources proposées, cependant notre expérience nous a montré que la pertinence prime sur la quantité d'information apportée à l'enseignant. Trop d'information tend à tuer l'information. Pour conclure, il apparait clairement que analyse évoquée avec vous n'est pas satisfaisante. Des modifications devront être effectuées.

Après recherche et modifications de l'algorithme les résultats sont un peu meilleurs :
Source 1 et 4 sont détectées et mise en relation : pour le reste certaines sources sont identifiées, mais ne sont pas reliées au document à tester.
Mail d'Urkund « Il s'avère que nous sommes J + 2 semaines de la mise en production de la nouvelle version d'Urkund, qu'une équipe des équipes travaille à l'implémentation de nouvelles fonctionnalités (Notamment les nouveaux formats)et que les mises à jour et problèmes de firewall ont perturbé voir stressée les différentes équipes techniques. Un concours de circonstances qui ont fait que l'équipe notre réponse n'a été ni rapide ni adéquate »

http://www.compilatio.net/
Contact : Jérome Savoie. Tel 04.79.25.75.45 jerome@sixdegres.fr

Payant ~~Gratuit~~. Langue : français. Installé sur serveur externe. Professeur envoie les copies ou demande à ses étudiants de la faire.

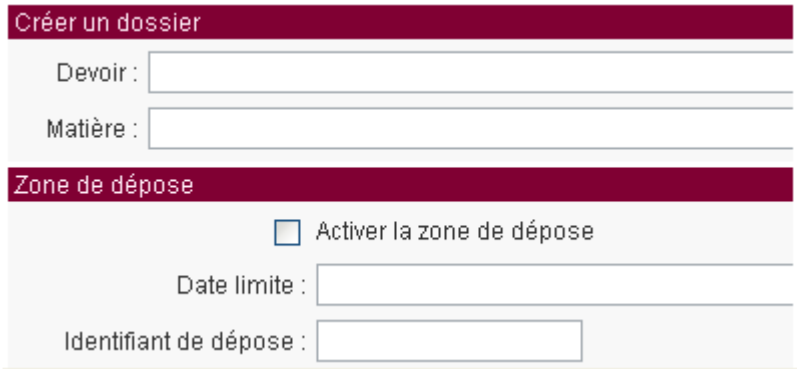
Code Test .
Login : xxxxxxx@wanadoo.fr Mot de passe : demoxxxxxxx

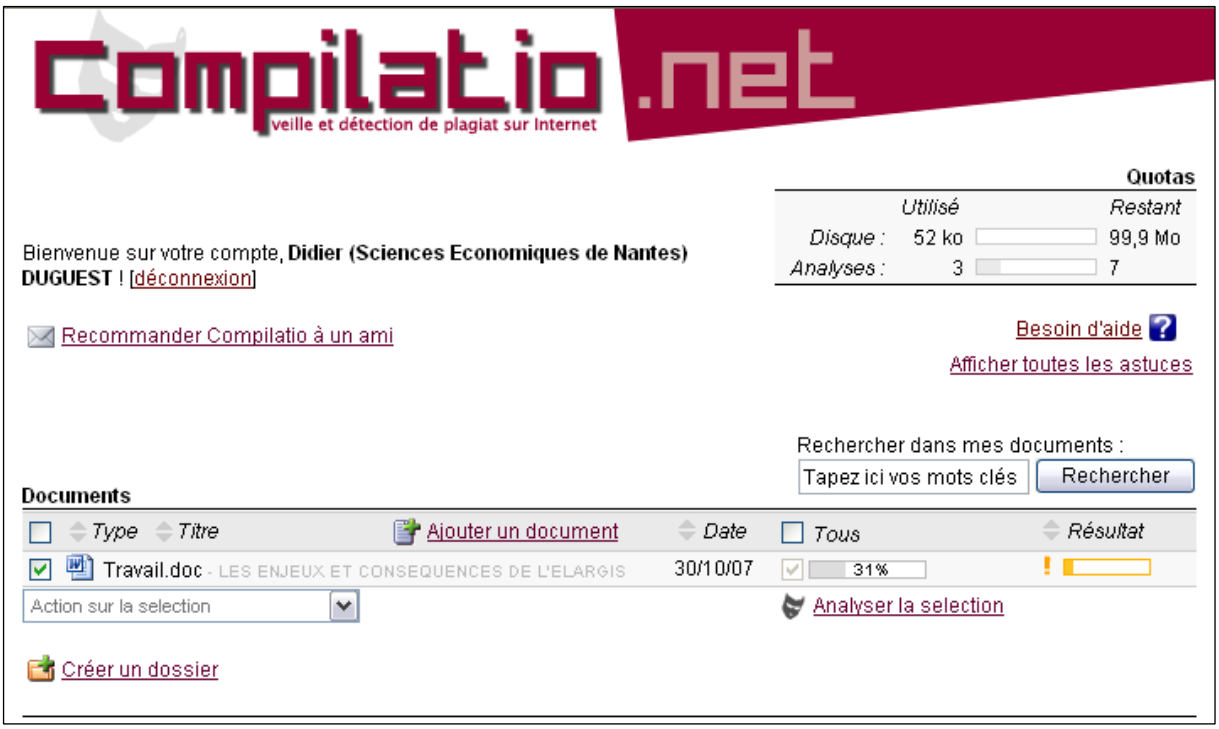
Fonctionnement simple.
La licence nous donne un login et un mot de passe.
Il suffit alors d'entrer les documents à analyser dans la base de données.
Puis après avoir sélectionner les documents que l'on souhaite analyser, on lance l'analyse (assez longue).
Le logiciel analyse tous types de documents (".doc", ".rtf", ".ppt", ".pdf", ".xls", ".txt", ".odt", ".html", etc.) et génère un rapport indiquant entre autre :
- le coefficient de plagiat du document
- les passages identifiés comme «copiés sur Internet»
- l'ensemble des sources de plagiats possibles.

Principe

Compilatio traite le document à analyser afin de ne conserver que son contenu textuel. Il extrait du document les passages dont l'analyse est pertinente, puis recherche tous les extraits identiques disponibles sur Internet ainsi que dans la base de documents du logiciel.
Tous les extraits retrouvés à l'identique sont considérés comme plagiés. Si un extrait est en fait une citation ou une phrase très courante, cet extrait peut être ignoré. Il ne sera plus pris en compte dans le calcul du coefficient de plagiat.
La première analyse d'un document s'effectue sur une centaine d'extraits du document (souvent une suite de 7 mots). Cette première analyse ne prend que quelques minutes. Pour certains très gros documents, plusieurs analyses sont nécessaires afin d'obtenir un résultat statistiquement significatif. Il recalcule automatiquement le coefficient d'authenticité du document sur la totalité des analyses effectuées.

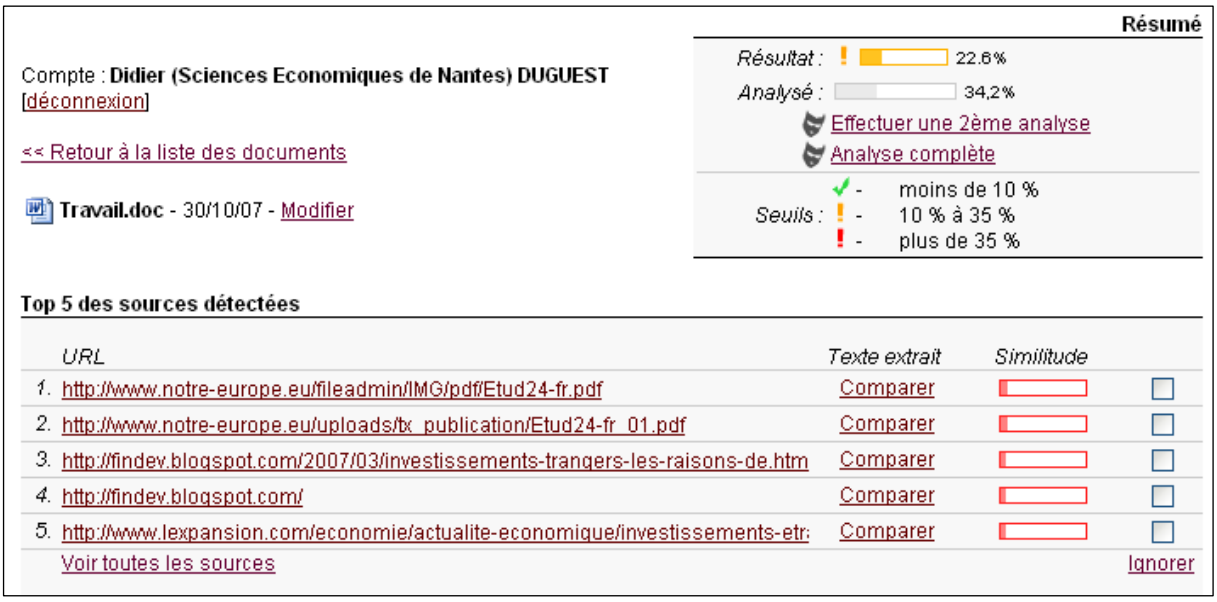
A noter la possibilité pour le professeur de créer un dossier, dans lequel les étudiants intègreront leurs documents. L'analyse se lance et les résultats sont envoyés au professeur.





A la fin de l'analyse, on obtient l'écran suivant, permettant de connaître l'état de plagiat du document.

Un clic sur la zone résultat, permet d'obtenir le détail :



Compilatio nous propose 325 sources sur ce document !

Vient ensuite le texte en noir avec les passages copiés en rouge.

Légende des couleurs
Texte retrouvé

Texte extrait du document

LES ENJEUX ET CONSEQUENCES DE L'ELARGISSEMENT DE L'UNION EUROPEENNE DE 15 A 25 PAR RAPPORT AUX IDE INTRA-EUROPEENS :

INTRODUCTION :

Depuis le debut de sa construction en 1951 avec le traite instituant la CECA (Communaute Europeenne du Charbon et de l'Acier), l'Union Europeenne (autrefois appelee Communaute Economique Europeenne) n'a cesse de s'agrandir, pour atteindre aujourd'hui le nombre de 27 pays membres (en 2007). Cet elargissement europeen, accompagne de quelques institutions (BCE...), a facilite la circulation d'IDE (investissement directs a l'etranger) intra europeens.

Les investissements directs a l'etranger consistent a injecter des fonds etrangers dans l'economie d'un pays dans le but de, soit y creer des entreprises nouvelles, soit de racheter une entreprise deja existant, tout ce processus devant aboutir a l'acquisition du pouvoir de **decision effectif dans la gestion de l'entreprise.** Par convention, une relation d'investissement direct est etablie des lors qu'un investisseur acquiert au moins 10 % **du capital social de l'entreprise investie.**

Un clic sur ces passages ouvre la source.

29. ... e. On montre que les investissements directs etrangers ont des effets plutot stabilisateurs dans les **pays en developpement alors que l'ouverture financiere** y a des effets plutot destabilisateurs. Au contraire l'ouverture financiere stabilise les economies ...

Ces mots ont été trouvés dans ces documents :	*Texte extrait*
1. http://enisse.kharroubi.free.fr/index2.html	Comparer
2. http://www.delta.ens.fr/abstracts/200426.html	Comparer

Résultat : 7 zones trouvées sur 9

Note 1 : non trouvée : le logiciel en trouve une partie dans un autre document. A noter que dans ce document figure l'intégralité de la phrase, mais la procédure Compilatio ne permet pas de l'identifier entièrement !
Zone 2 : OK
Zone 3 : OK avec une erreur en fin de phrase.
Zone 4 : OK
Zone 5 : OK
Zone 6 : OK
Zone 7 : Phrase modifiée : pluriels mis au singulier. La partie de la phrase avec les fautes n'est pas retrouvée. Le reste de la phrase (non modifiée) est retrouvée à la bonne source.
Zone 8 : OK
Zone 9 : OK

A noter que de nombreux passages sont considérés comme plagiés alors qu'ils sont simplement des parties de phrase, confinant parfois au ridicule. Ainsi la suite de 7 mots 'de vendre des produits à bas prix' contenu dans notre document, est considérée comme plagiée et 6 références Internet sont présentées comme sources potentielles !

En consequence, l'integration au Marche Unique fera disparaitre cette barriere commerciale entre nouveaux adherents et anciens membres : Ces derniers ne pourront plus engager de procedures anti-dumping (fait **de vendre des produits a bas prix dans un autre pays du fait de** faibles contraintes sociales, fiscales, ou environnementales) **a l'egard des PECO une fois que** ceux-ci seront membres de l'UE.

17. ... dherents et anciens membres : Ces derniers ne pourront plus engager de procedures anti-dumping (fait **de vendre des produits a bas prix** dans un autre pays du fait de faibles contraintes sociales, fiscales, ou environnementales) a l'egar ...

Ces mots ont été trouvés dans ces documents :	*Texte extrait*
1. http://www.lefigaro.fr/eco-entreprises/20061113.FIG000000255_les_epiciers_de	Comparer
2. http://lematindabidjan.com/visual_article.php?num_actualite=6646	Comparer
3. http://www.3s.hmg.inpg.fr/ci/doc/phd-thesis/19991204-Guy_Prudhomme.pdf	Comparer
4. http://www.lematindabidjan.com/visual_article.php?num_actualite=6646	Comparer
5. http://www.tourmag.com/index.php?action=article&id_article=212704	Comparer
6. http://forum-des-jeux.actifforum.com/Questions-Reponses-f29/produits-de-base-	Comparer

Comme indiqué ci-dessus, ce logiciel sélectionne un certain nombre de suites de 7 mots consécutifs, formant une phrase et recherche celle-ci sur Internet. Si ce texte est retrouvé, l'original est alors comparé avec la copie étudiante, afin de savoir si la plagiat se limite à cette succession de 7 mots ou est plus important. Cela entraîne une avalanche de sources, diminuant la lisibilité des résultats.

Ainsi la phrase

La solidarité intra-européenne est alors renforcée.

On montre que les investissements directs étrangers ont des effets plutôt stabilisateurs dans les pays en développement alors que l'ouverture financière y a des effets plutôt déstabilisateurs. Au contraire l'ouverture financière stabilise les économies développées. [4]

Est découpée en 5 zones de 7 mots, avec les mêmes sources :

27. ... faciles du fait de la suppression des entraves. La solidarite intra-europeenne est alors renforcee. **On montre que les investissements directs etrangers** ont des effets plutot stabilisateurs dans les pays en developpement alors que l'ouverture financiere ...

Ces mots ont été trouvés dans ces documents :	*Texte extrait*
1. http://enisse.kharroubi.free.fr/index2.html	Comparer
2. http://www.delta.ens.fr/abstracts/200426.html	Comparer

28. ... solidarite intra-europeenne est alors renforcee. On montre que les investissements directs etrangers **ont des effets plutot stabilisateurs dans les** pays en developpement alors que l'ouverture financiere y a des effets plutot destabilisateurs. Au co ...

Ces mots ont été trouvés dans ces documents :	*Texte extrait*
1. http://enisse.kharroubi.free.fr/index2.html	Comparer
2. http://www.delta.ens.fr/abstracts/200426.html	Comparer

29. ... e. On montre que les investissements directs etrangers ont des effets plutot stabilisateurs dans les **pays en developpement alors que l'ouverture financiere** y a des effets plutot destabilisateurs. Au contraire l'ouverture financiere stabilise les economies ...

Ces mots ont été trouvés dans ces documents :	*Texte extrait*
1. http://enisse.kharroubi.free.fr/index2.html	Comparer
2. http://www.delta.ens.fr/abstracts/200426.html	Comparer

30. ... ont des effets plutot stabilisateurs dans les pays en developpement alors que l'ouverture financiere **y a des effets plutot destabilisateurs.** Au contraire l'ouverture financiere stabilise les economies developpees. 2.2 : Les effets des IDE po ...

Coût

La licence peut être solo ou groupée (licence globale pour l'ensemble des enseignants d'un établissement.

Compte Solo : tarifs TTC. Espace de stockage 100 Mo.
Offert : 5 analyses 149,00 €
Remarques :
- engagement minimum 12 mois.
- analyses offertes non reportables.

Compte utilisateur

Les documents chargés par les utilisateurs sont stockés sur le serveur de Compilatio. Chaque compte Solo bénéficie d'un espace de stockage permettant de stocker environ 200 documents.

Crédits d'analyse

Une fois stockés, ces documents peuvent être analysés. Chaque utilisateur possède un crédit d'analyse qui détermine le nombre d'analyse qu'il peut effectuer.
Chaque compte Solo est livré avec 5 analyses gratuites. Ainsi, Compilatio est immédiatement opérationnel, et il est possible de mener une action crédible en effectuant quelques analyses parmi les travaux qui semblent les plus "douteux".

Type d'offre

Solo : *utilisateur unique.*

Pour toute autre formule (multi-utilisateurs, intégrations au plate-formes pédagogiques, etc.), contactez-nous :
contact (arobase) compilato.net

Durée de l'abonnement

12 mois à compter d'aujourd'hui

Votre abonnement prendra effet aujourd'hui, le 30 octobre 2007, et s'étendra jusqu'au 31 septembre 2008 inclus.

	Tarif HT	*Tarif TTC*
Espace de stockage		
• *100 Mo d'espace disque* • *5 analyses offertes*	124,59 €	149,00 €
Crédits d'analyses		
500 analyses - supprimer **Offre promotionnelle** : +50% soit **250** analyses offertes pour votre premier abonnemement	417,98 €	499,90 €
Ajouter d'autres analyses		
Total	**542,56 €**	**648,90 €**

http://www.turnitin.com/plagiarism_i18n.html?lang=fr
Contact : Samantha Allan Tel: + 44 (0) 191 243 8442. sam.allan@northumbrialearning.co.uk

Payant ~~Gratuit~~. Langue : français. Installé sur serveur externe. Professeur demande à ses étudiants d'envoyer les copies.

Email Address: xxxxxx@wanadoo.fr
Password: xxxxxxxx

Prévention contre le plagiat

Reconnu dans le monde entier comme la norme pour la prévention contre le plagiat en ligne, Turnitin aide les enseignants et les étudiants à obtenir tous les avantages du potentiel éducatif d'Internet.

Une recherche complète

Chaque devoir envoyé est retourné sous la forme d'un Rapport d'Originalité personnalisé. Les résultats sont basés sur des recherches complètes de milliards de pages d'exemples actuels et archivés de l'Internet, sur les millions de devoirs d'étudiants auparavant envoyés à Turnitin, et sur les bases de données commerciales d'articles de journaux et de périodiques.

Technologie de propriété par recoupement

La technologie de recherche la plus avancée de l'industrie qui vérifie des devoirs par rapport à nos copies internes du contenu Internet actuel et archivé et par rapport à notre base de données de propriété de millions de devoirs d'étudiants précédemment envoyés.

Tout simplement puissante

Contrairement à certains autres systèmes de prévention contre le plagiat, Turnitin n'exige aucune installation ou maintenance de logiciel supplémentaire. Parce que Turnitin est entièrement basé sur l'Internet, la compatibilité entre différents ordinateurs et systèmes d'exploitation n'est jamais un problème.

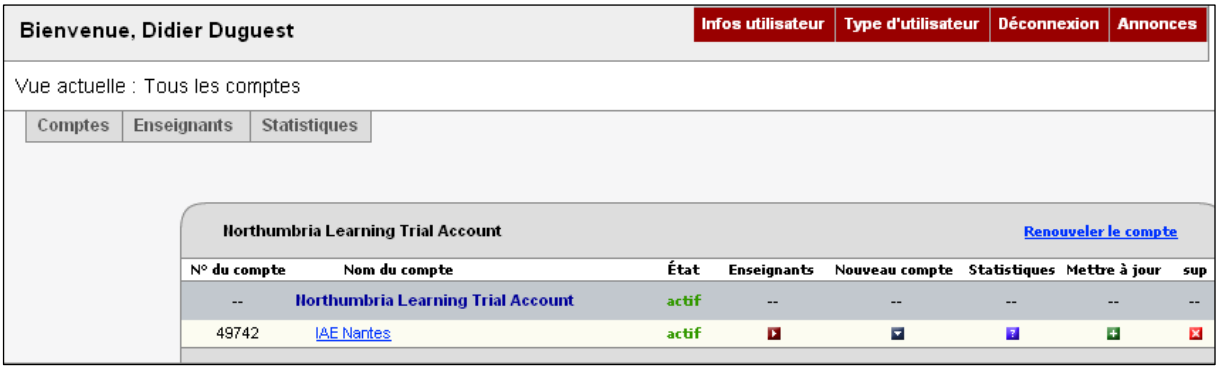
De façon plus importante, l'aide au démarrage de la prévention contre le plagiat est rapide et facile. Notre matériel de support sérieux, notre conception d'interface intuitive, nos programmes de formation complets et notre dévouement au support clients de qualité garantissent que l'utilisation de Turnitin est une expérience constamment gratifiante.

Un rapport détaillé

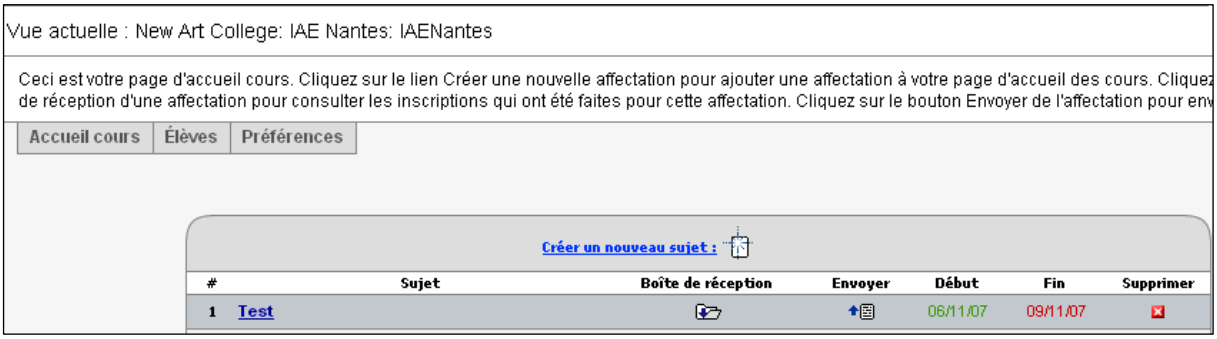
Toute similitude non découverte entre des devoirs envoyés et le matériel source est exposée en détail dans un format intuitif et non équivoque, permettant aux enseignants de passer plus de temps à traiter les causes du plagiat plutôt que de les rechercher.

La prévention contre le plagiat de Turnitin est souvent si réussie que les institutions qui utilisent notre système à une grande échelle voient des taux mesurables de plagiat atteindre pratiquement zéro.

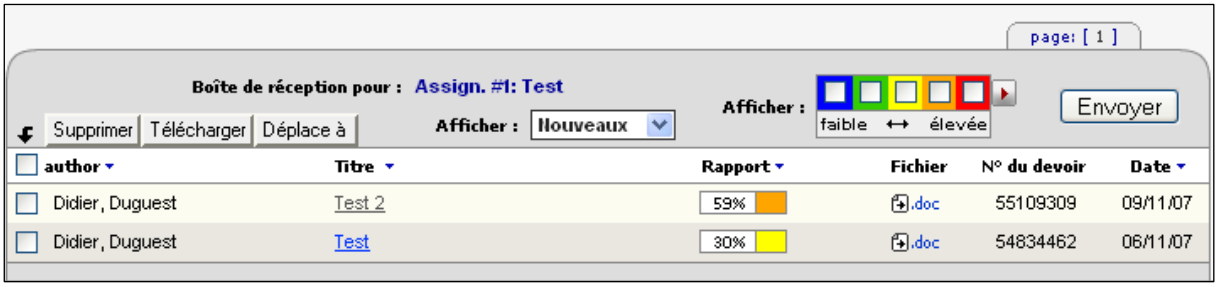
La mise en place du compte est assez longue.



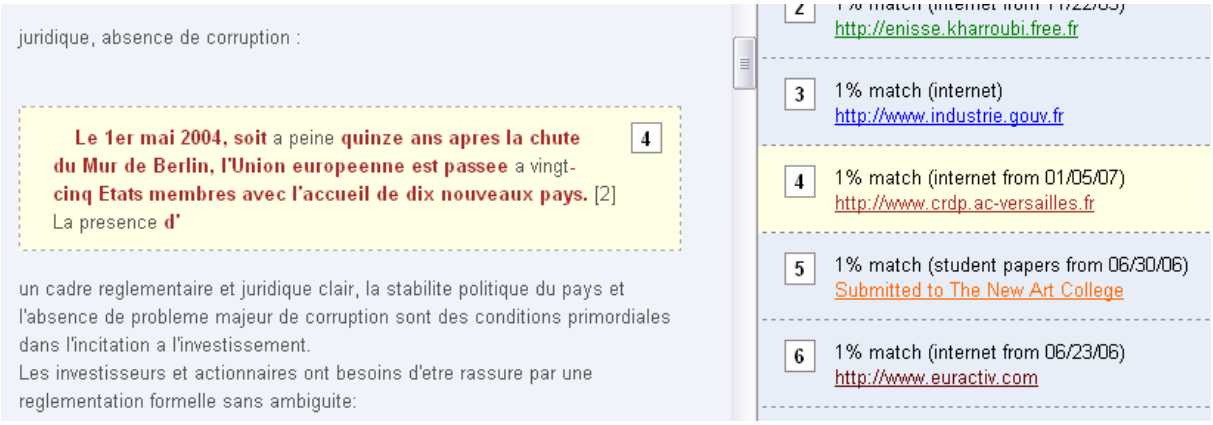
Il est alors possible de créer des comptes spécifiques à chaque enseignant. L'enseignant communiquera alors les identifiants et date limite de remise aux étudiants.
Lors de la remise des travaux cet écran apparaît :



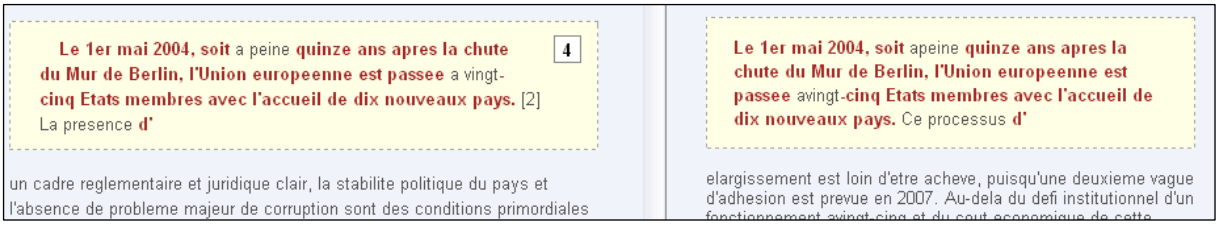
Résultat
Les résultats donnent, dans un premier temps le pourcentage de plagiat (ici 30%) :



Lorsqu'un passage est considéré comme plagié, il apparaît en encadré avec un numéro à gauche de l'écran avec la correspondance à droite.



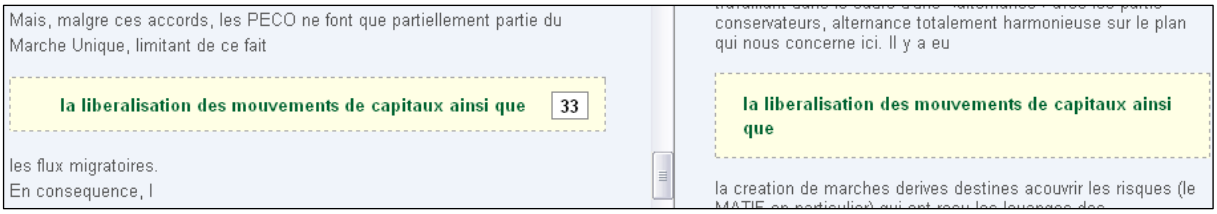
En cliquant sur la zone on ouvre à droite le document source, permettant ainsi d'évaluer la réalité du plagiat de façon directe.



Résultat : 6 zones trouvées sur 9

Zone 1 : OK. Source différente due aux copies de copies sur Internet !
Zone 2 : OK. Même remarque
Zone 3 : OK.
Zone 4 : OK.
Zone 5 : OK.
Zone 6 : Non trouvée.
Zone 7 : Phrase modifiée : pluriels mis au singulier. Non trouvée.
Zone 8 : Ok, mais source différente, retrouvée dans Google en recherche élargie.
Zone 9 : Non trouvée.

Comme pour Compilatio, on retrouve des zones considérées comme plagiées (38 sources trouvées !) alors que ce sont simplement des coïncidences, par exemple



Prix :
The price for a Turnitin Licence is £1,000 for the licence, plus an extra £0.54 GBP per student. If your student enrolment is 500, the total cost would be around £1,270.00 based over a 12 month period. This cost allows for an unlimited number of submissions and tutor enrolment per year, helpdesk and email support and a free plug-in available to integrate with Moodle. We also have departmental licences available if this interests you. The cost for a departmental licence is £300 for the licence plus 0.77 per student.

http://www.ephorus.nl/
Contact Steven De Weirdt Tel +32-(0)498.110.258 steven.de.weirdt@ephorus.com

Payant ~~Gratuit~~. Langue : français. Installé sur serveur externe. Professeur envoie les copies ou demande à ses étudiants de la faire.

Login : xxxxxxx@wanadoo.fr

Mot de passe : xxxxxxx

Fonctionnement d'Ephorus

1) **Préparation du document**

Le document (.doc, .pdf, .txt, .docx, .odt, .sxw, .html) est converti en fichier texte (txt) sans mise en page ou photos.

Le texte est découpé en morceaux (nommées les 'queries'). Ces parties sont constituées au minimum de 7 mots de 3 caractères minimum, en dehors des codes de contrôle.
Chaque morceau est contrôlé par le moteur de recherche sur la Base de données Internet et la base de données de l'établissement d'enseignement (et éventuellement des autres bases de données).

2) **Création du rapport**

Les sources trouvées sur les sites Web et les documents autres avec des pourcentages de concordance importants sont alors téléchargés, puis comparés au document source.

Il n'est pas indispensable que chaque mot soit identique pour conclure au plagiat 'pour éviter les introduction de fautes d'orthographes et synonymes). L'enseignant peut, dans le rapport définir la rigueur de contrôle (stricte, standard, adaptable).

Sur base de ces comparaisons, le rapport de plagiat est établi : le document original est mis en parallèle des sources pour une comparaison plus facile.

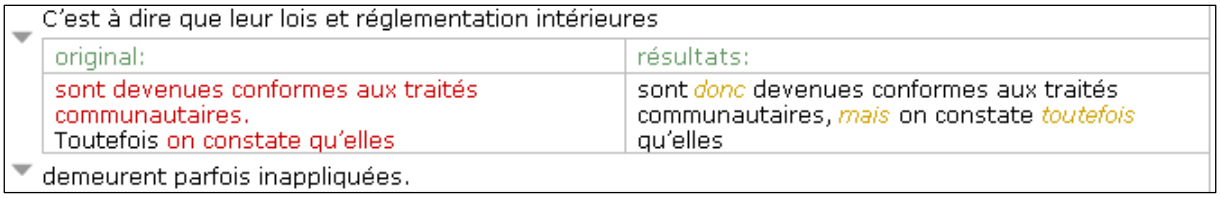
3) **Transmission du rapport**

Le rapport est envoyé vers la page personnelle de l'enseignant sur l'application ou sur votre plateforme d'apprentissage électronique.

ephorus
Mes documents
Télécharger
Mon compte
Assistance
Contact
Connecté en tant que: Duguest Didier (fermer la session)
Télécharger
Télécharger document
Télécharger zip
Télécharger document
Après le téléchargement, le document apparaîtra dans votre boîte e-mail.
Titre du document
Test
Rechercher le document
G:\Plagiat\Travail.
Parcourir...
Options de téléchargement:
Visible – Effectuer une vérification (par défaut)
Visible – Ne pas effectuer de vérification
Invisible – Effectuer une vérification
Télécharger

ephorus
Mes documents
Télécharger
Mon compte
Assistance
Contact
Connecté en tant que: Duguest Didier (fermer la session)
Téléchargements manuels
Score
Rapport
Date
12%
Test
06/11/2007 15:21
Page 1

Pour lire les résultats, il suffit de cliquer sur la source retrouvée : le logiciel détaille alors, dans le document d'origine, les zones de similitudes.
Possibilité intéressante : lorsqu'une zone plagiée est repérée, le logiciel propose en vis-à-vis la copie et la version originale, ce qui permet également de voir les modifications apportées par l'étudiant et déterminer la 'qualité' du plagiat :



Résultat : 3 zones trouvées sur 9

Zone 1 : non trouvée.
Zone 2 : non trouvée.
Zone 3 : non trouvée.
Zone 4 : OK
Zone 5 : OK
Zone 6 : non trouvée.
Zone 7 : Phrase modifiée : pluriels mis au singulier. non trouvée.
Zone 8 : non trouvée.
Zone 9 : OK

Premières conclusions

Après ce premier test quatre logiciels sortent du lot : Compilation, Ephorus, Urkund et Turnitin. Ce qui ne semble pas anormal puis que tous, en dehors de Compilatio, logiciel français ne participant généralement pas aux tests, arrivent en tête des tests d'évaluation des logiciels de plagiat.
Ceci étant si les résultats sont plutôt positifs, il reste des questions sans réponses :

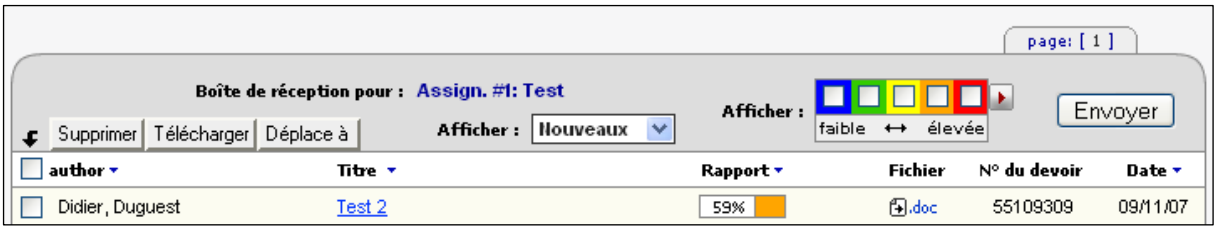
- Compilation est performant mais fournit une liste de sources indigestes, rendant l'ensemble illisible. Par ailleurs son processus de recherche amène à considérer comme plagiat des phrases qui ne sont pas plagiées.
- Urkund, généralement bien placé dans les tests, nous donne ici des résultats plus que moyens. L'équipe technique est sur le pied de guerre pour résoudre ce problème.
- Turnitin donne de très bons résultats, mais la question de savoir pourquoi la dernière phrase plagiée (numéro 9) reste sans vraie réponse. Les questions posées au responsable génèrent des réponses souvent a minima.
- Ephorus donne des résultats qui peuvent, en première lecture, paraître moyens, mais fournit une explication plausible : la recherche se fait sur des suites de 130 caractères en moyenne, ce qui explique que les sources 1, 2, 3 et 6 ne soient pas trouvées. Par ailleurs le responsable de cette société insiste sur le fait qu'il est à l'écoute des critiques pour améliorer son produit.

Pour poursuivre cette étude j'ai créé un autre document (test2.doc) dans le but de vérifier la capacité des logiciels à détecter des sources assez peu répandues sur Internet ainsi que la possibilité de contourner le fait que les étudiants puissent introduire des fautes d'orthographe et des synonymes à l'intérieur des zones plagiées, pour contourner les recherches anti plagiat.
L'intégralité du document de base est une partie d'annexe d'un sujet d'examen BTS Action co datant de 2000, se trouvant dans les annales des éditions Foucher. J'ai collé 6 zones plagiées, en provenance d'Internet, 5 de langue française et 1 de langue anglaise (en rouge dans le document). J'ai introduit dans ces zones plagiées des fautes d'orthographes (pluriels mis au singulier) et quelques synonymes (augmentation / croissance) : ces modifications sont en bleu.
J'ai alors soumis ce nouveau document à la sagacité des 4 logiciels nommés ci-dessus.
Les 6 sources sont les suivantes :
http://www.terra-economica.info/Le-lavage-de-voiture,2500.html
http://core.asso.free.fr/kakapo/eau.htm
http://www.reg-grenoble.fr/pdf/ecogal.pdf
http://www.maisoneco.com/article27.html
http://www.liberation.fr/php/pages/pageReactionsList.php?rubId=15&docId=195521
http://www.ehow.com/how_1328_wash-car.html

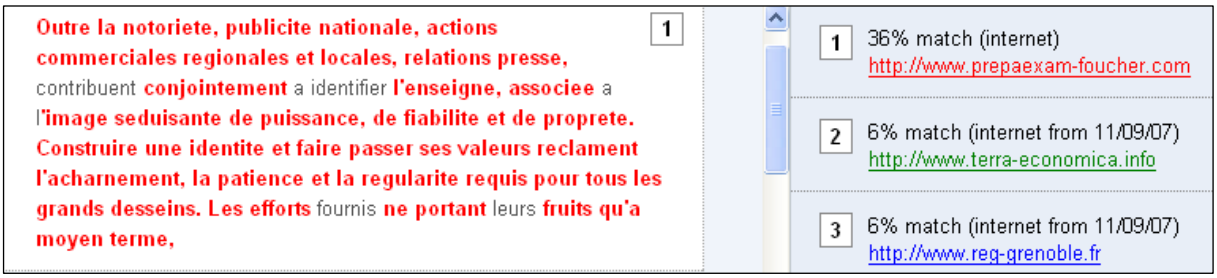
ainsi que la source du document d'origine :
http://www.prepaexam-foucher.com

Le texte comporte 922 mots, 389 sont plagiés soit 42 %

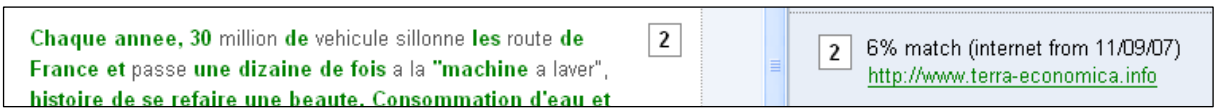
59 % du document est considéré comme plagié.



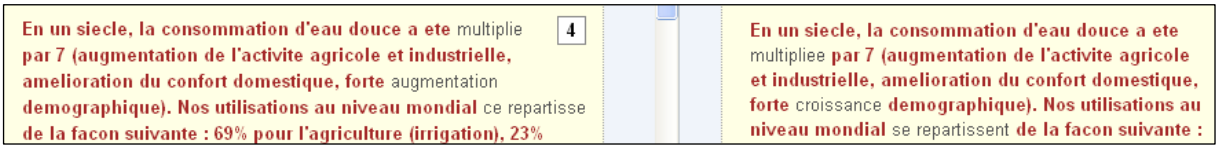
Turnitin retrouve la source d'origine du document :



Les fautes d'orthographe sont contournées :



Ainsi que les utilisations de synonymes :



Résultat : 4 zones plagiées détectées sur 6.

Document d'origine : OK
Zone 1 : OK
Zone 2 : OK
Zone 3 : OK
Zone 4 : Non trouvée. Comporte 3 synonymes.
Zone 5 : Non trouvée. Article de libération.
Zone 6 : OK, mais avec une Zone différente.

ephorus

2 % du document est considéré comme plagié, avec 1 seule source !!!

Téléchargements manuels

Score	Rapport	Date
2%	Test 2	09/11/2007 19:33

Curieusement seule une partie de la zone 3 est identifiée, mais partiellement seulement :

Connecté en tant que: Duguest Didier (fermer la session) Rapports: résumé / détaillé / version imprim

La surface de la terre est recouverte d'eau à 70 %, mais

original:	résultats:
97.5 % de cette eau est salée. Les 2.5 % restants sont de l'eau douce, dont près des trois quarts sont gelés, ce qui laisse moins de 1 % de	97.5% de cette eau est *sale*, les 2.5% restants sont de l'eau douce, dont *prs* des trois quarts sont *gels*, ce qui laisse moins de 1% de

Résultat : 1 zone plagiée détectée sur 6.

Document d'origine : Non trouvée.
Zone 1 : Non trouvée. Site info de terra Economica.
Zone 2 : Non trouvée. Site perso.
Zone 3 : OK, partiellement
Zone 4 : Non trouvée. Article de Maison Eco
Zone 5 : Non trouvée. Article récent de Libération
Zone 6 : Non trouvée. Site anglais

J'ai contacté le service technique afin d'obtenir une explication sur la faiblesse de ces résultats. Le technicien a repris mon document, l'a modifié (en prenant des zones plagiées plus grandes que le miennes) et soumis de nouveau. Ses résultats sont meilleurs que les miens : probablement du au fait que les zones plagiées dans son document sont plus importantes que dans le mien : si nous choisissons cette solution il faudra demander aux étudiants de veiller à élargir autant que faire se peut les zones copiées !
Ci-dessous résultat et commentaire du technicien.

- Concernant le premier site web : http://www.terra-economica.info/Une-voiture-creee-de-toutes-pieces,3240.html?var_recherche=voiture

  Ce texte n'est pas trouvé… peut être parce que un grand partie du site et protége avec un mot de passe. Ce n'est pas possible de lire tous les articles sans registrer. Si vous placer le texte sur Google, le moteur de recherche ne trouve pas le texte non plus.

- Deuxième site web : http://core.asso.free.fr/kakapo/eau.htm

  Maintenant le plagiat est trouve, ainsi que les fautes d'orthographes et synonymes. Mon partie et plus grand.

- Concernant le site web numéro 3 : http://www.reg-grenoble.fr/pdf/ecogal.pdf

  Le PDF n'est pas trouvé. Je n'ais pas d'explication. Pourtant Ephorus trouve des certaines PDF.

- Site Web Numéro 4 : http://www.maisoneco.com/article27.html

  Le plagiat est trouvé ainsi que le changement « contaminations ». Mon partie et plus grand.

- Site Web 5: http://www.liberation.fr/php/pages/pageReactionsList.php?rubId=15&docId=195521

  Le plagiat n'est pas trouvé. Peut être parce que si on fait un copy-paste de cette page, les résultats sont dans des cadres (frames). C'est un forum avec des petits messages en cadres.

- Site Web 6 : http://www.ehow.com/how_136964_keep-car-clean.html

  Le plagiat est trouve, ainsi que les fautes et synonymes. Mon partie et plus grand.

Je suis désolé mais je ne peut pas vous donner des explications exacte pour laquelle mon test et mieux que le votre. Le seul chose c'est que moi j'ai prix des parties de texte plus grand que le votre, et maintenant les résultats sont mieux, mais c'est vrai… pas tous le plagiat est trouve.

34 % du document est considéré comme plagié, dans 44 sources !

Résumé

Compte : **Didier (Sciences Economiques de Nantes) DUGUEST**
[**deconnexion**]

Résultat : ! 34.0%
Analysé : 100,0%

Compilatio ne retrouve pas la source d'origine du document.

Nous retrouvons l'inconvénient rédhibitoire des conséquences de l'algorithme de recherche : la source 1 est trouvée, mais partiellement et découpée en 6 morceaux : la première partie de la phrase comportant des fautes d'orthographe n'est pas identifiée et le reste est découpée par zones de 7 mots :

6. ... ement, pres de 35 millions de metres cube d'eau sont employes chaque annee au lavage de voiture dont **plus de la moitie au domicile des automobilistes.** A travers la cohesion du reseau et son homogeneite, Elephant Bleu credibilisera son offre en menant ...

Ces mots ont été trouvés dans ces documents : *Texte extrait*
1. http://www.terra-economica.info/Le-lavage-de-voiture,2500.html **Comparer**

Cet algorithme comporte toujours le risque de faux plagiats :

themes de communication definis dans la campagne 1998 ont tous ete declines au niveau national. Traites differemment regionalement et localement, l'impact des operations **a ete plus ou moins satisfaisant.**

Certains résultats viennent contredire les affirmations du créateur du logiciel, selon lequel, lorsqu'une source est trouvée de façon aléatoire, celle-ci est approfondie afin de déterminer le degré de plagiat : ici malgré une découverte de plagiat, des zones plagiées ne sont pas analysées :

La recuperation d'eau est un geste citoyen a portee de tous. **En plus de vous faire economiser le prix de l'eau, c'est un geste de preservation de** cette denree rarissime qu'est l'eau et cela participe a la prevention des inondations. **Pour l'arrosage du jardin, tout le monde peut** installer une cuve exterieure au bout d'une gouttiere. **On en trouve dans le commerce des petites** cuves de 50 a 300 litres, generalement en plastique.

Pour l'arrosage du jardin, tout le monde peut installer une cuve externe au bout d'une gouttière. **On en trouve dans le commerce des petites** cuves de 50 à 300 litres, en général en plastique.

Pour aller plus loin, il est nécessaire d'enterrer une cuve, béton de préférence (elle

Globalement. 5 zones plagiées détectées sur 6.

Document d'origine : Non détecté.
Zone 1 : OK, mais sans la détection des zones mal orthographiées.
Zone 2 : OK, mais avec le même problème : incapacité de détecter les zones avec synonymes ou mal orthographiées.
Zone 3 : OK mais détection partielle, avec une Zone différente. (en fait la Zone est une copie modifiée de l'original !)
Zone 4 : Détection partielle. Comporte 3 synonymes.
Zone 5 : OK
Zone 6 : OK

Suite à de nouveaux problèmes pour déposer le document sur le site, l'analyse a été reportée de plusieurs jours, pour finir par donner un résultat plus que moyen.
Les sources identifiées sont incomplètes, et surtout ne pointent pas vers les passages du document source !

Liste des sources

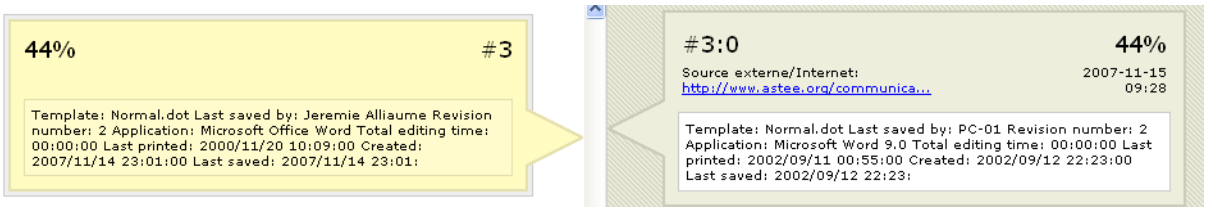
Type de source	URL/Nom de fichier	Mémorisé en cache
Source externe/Internet	http://www.un.int/france/documents_francais/020902...	2007-11-15 09:28
Source externe/Internet	http://www.monde-arabe.u-3mrs.fr/MACLOTc.html	2007-11-15 09:28
Source externe/Internet	http://www.diplomatie.gouv.fr/fr/article-imprim.ph...	2007-11-15 09:28
Source externe/Internet	http://www.astee.org/communication/cai/fichiers/2_...	2007-11-15 09:28
Source externe/Internet	http://www.terra-economica.info/Le-lavage-de-voitu...	2007-11-15 09:28

Résultat à cette adresse :
http://a2.urkund.se/view/747437-791345-623038

Résultat : 1 zone plagiée détectée sur 6.

Document d'origine : Non trouvé
Zone 1 : OK
Zone 2 : Non trouvée. Site perso.
Zone 3 : Seule la dernière partie de la phrase est détectée.
Zone 4 : Non trouvée. Article de Maison Eco
Zone 5 : Non trouvée. Article récent de Libération
Zone 6 : Non trouvée. Site anglais

Une source est détectée, sans rapport avec l'étude !



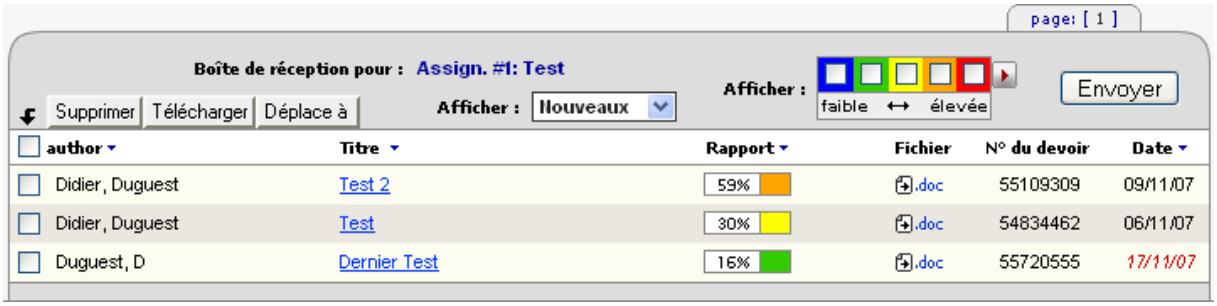
Donc toujours d'importants problèmes, soit disant dus à la mise en place d'un nouvel algorithme ! Peut être vrai, mais impossible de juger de la vraie valeur de cette solution.

Troisième test

Pour terminer cette première analyse j'ai fait un troisième test. Partant d'un article paru sur le site de l'Humanité en décembre 1998 sur un grand cuisinier de la région de St Malo, Olivier Rollinger :
http://www.humanite.fr/popup_imprimer.html?id_article=431084
j'ai ajouté des sources plus ou moins récentes, prises sur Internet, dans des formats html, pdf ou sur des blogs en langue anglaise. L'article (LaRolle.doc) comporte 9 sources plagiées :
http://lemondeenmarche.uniterre.com/28036/Mardi+23+octobre+Cancale.html
http://www.jedecouvrelafrance.com/f-3616.ille-et-vilaine-huitres-cancale.html
http://www.lexpansion.com/economie/actualite-economique/consommation-de-vin-les-etats-unis-devant-la-france-en-2008_107889.html
http://www.lemonde.fr/web/article/0,1-0@2-3234,36-977607@51-959139,0.html
http://www.oboulo.com/search.php?S=STR&T=&q=&L=fr&N=&E=&start=80
http://www.forumducommerce.org/news/fullstory.php/aid/272.html
http://dsp-psd.pwgsc.gc.ca/Collection/E12-9-19-10F.pdf
http://fr.wikipedia.org/wiki/Calvados_(alcool)
http://www.vintnersgroup.com/Food%20blog/Olivier%20Roellinger/index.html

Afin de les identifier facilement, ces zones sont en rouges dans le texte avec des notes en bas de page contenant les sources.
Les zones modifiées représentent 546 mots sur un total de 2042 soit 27 %. Il est par ailleurs possible de considérer que ce document est plagié à 100 %, puisque l'intégralité du contenu provient d'Internet.
J'ai enfin modifié quelques unes des zones plagiées (fautes d'orthographes, synonymes) pour tester l'efficacité sur ces points. Les zones modifiées sont en bleu.

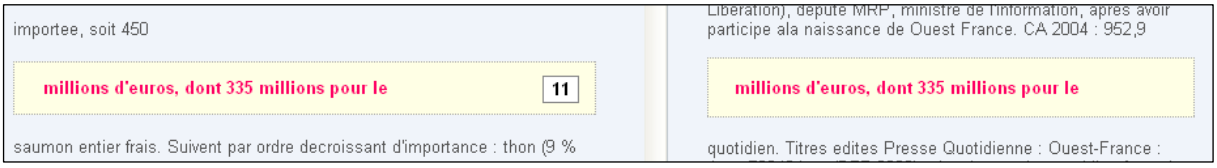
16 % du document est considéré comme plagié dans 16 sources.



Turnitin ne retrouve pas la source d'origine du document.

Les fautes d'orthographe sont contournées, ainsi que les utilisations de synonymes, lorsque la source est identifiée :

Attention faux plagiats : on confond le saumon et Ouest France



Résultat : 3 zones plagiées détectées sur 9

Document d'origine : Non trouvé
Zone 1 : Non trouvée. Blog récent.
Zone 2 : OK
Zone 3 : OK
Zone 4 : Non trouvée. Article du Monde
Zone 5 : Non trouvée. Site Oboulo.
Zone 6 : OK.
Zone 7 : Non trouvée. Faux plagiat.
Zone 8 : Non trouvée. Wikipédia.
Zone 9 : Non trouvée. Blog récent.

ephorus

14 % du document est considéré comme plagié avec 4 sources différentes.

Connecté en tant que: Duguest Didier (fermer la session)

Téléchargements manuels

Score	Rapport	Date
14%	Test3	18/11/2007 10:46

document original	
	Test3
14% score total	
3%	http://www.forumducommerce.org/news/fullstory.php/aid/272/Les_%E9pices.html
4%	http://fr.wikipedia.org/wiki/Calva
2%	http://www.jedecouvrelafrance.com/f-3616.ille-et-vilaine-huitres-cancale.ht...
3%	http://www.oboulo.com/strategie-44-expose.html

Confirmation du fait que Ephorus ne reconnaît pas les accents :

original:	résultats:
certaines atteigne même le calibre "Pied de Cheval". Les parcs s'étendent sur plus de 350 hectares. C'est une huître à chair ferme, qui nécessite 3 à 4 années d'élevage, qui se mange aussi bien crue que cuisinée. Elle a une saveur particulière de noisette et un	certaines *atteignent* le calibre "Pied de Cheval". Les parcs *s'tendent* sur plus de 350 hectares. C'est une *hutre* chair ferme, qui *ncessite* 3 4 *annes* qui se mange aussi bien crue que *cuisine.* Elle a une saveur *particulire* de noisette et un

Résultat : 4 zones plagiées détectées sur 9

Document d'origine : Non trouvé
Zone 1 : Non trouvée. Blog récent.
Zone 2 : OK
Zone 3 : Non trouvée. Article de l'Expansion.
Zone 4 : Non trouvée. Article du Monde
Zone 5 : OK.
Zone 6 : OK.
Zone 7 : Non trouvée. Fichier pdf
Zone 8 : OK.
Zone 9 : Non trouvée. Blog récent.

30 % du document est considéré comme plagié, dans 148 sources !
Compilation identifie la source d'origine et indique un taux de plagiat de 81 %. Possibilité d'ignorer cette source, pour une information plus représentative.

Compte : [deconnexion]

<< Retour à la liste des documents

LaRolle.doc - 17/11/2007 - Modifier

Résultat : ! 29.6%
Analysé : 42,9%
Effectuer une 2ème analyse
Analyse complète

Nous retrouvons de nouveau le risque de faux plagiats (Noter l'inconvénient de ne pas avoir la source en couleur)

et en le masquant de sauce blanche. Soyons simple. Une echalote grise et une pointe d'ail tombees au beurre. Aller-retour dans la poele et on remet dans la coquille. **On peut meme les preparer a l'avance.** Au dernier moment, il n'y a plus qu'a les glisser sous le gril. Dans le verre, toujours le vouvray sec, ou bien un savennieres.	On sert les mousses et autres tartinades avec des tortilla campagne grillé. On peut même les préparer à l'avance, 2-3 jours au réfrigérateur, rangées dans des récipients he

Les fautes d'orthographes et les synonymes ne sont pas détectés :

Placee sur des tables en fer surelevees, l'huitre est decouverte et recouverte a chaque maree, lui assurant ainsi une oxygenation optimale et une apport en plancton fort interessant... certaines atteigne meme le calibre Pied de Cheval . **Les parcs s'etendent sur**

Ce qui explique peut être que les zones 3 et 7, dans lesquelles synonymes et fautes d'orthographe ont été introduit, n'aient pas été détectées**.**

contrarierait nullement ce gourmet. Alors que la consommation de vin ne cesse de diminuer en France, elle progesse dans le reste du monde. Elle devrait ainsi augmenter de 6,8% d'ici 2008, selon une etude du cabinet britannique ISWR/DGR publiee jeudi par le Salon mondial 2005 des vins et spiriteux, Vinexpo. Amples, gras, frais, puissants, avec ce merveilleux bouquet ou se marient le miel, le tilleul,

Résultat : 7 zones plagiées détectées sur 9

Document d'origine : Trouvé
Zone 1 : OK.
Zone 2 : OK
Zone 3 : Non trouvée. Article de l'Expansion.
Zone 4 : OK
Zone 5 : OK
Zone 6 : OK.
Zone 7 : Non trouvée. Fichier pdf.
Zone 8 : OK
Zone 9 : OK

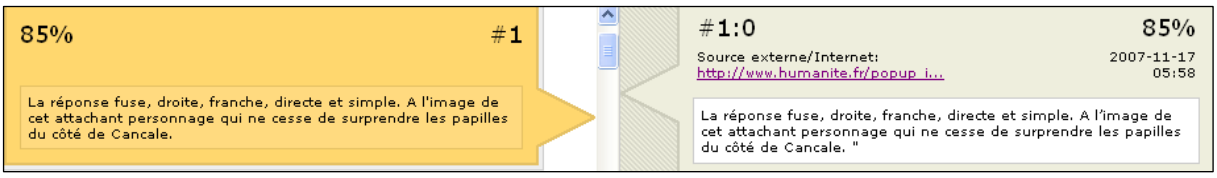
Enfin Urkund fonctionne !
67 % du document est considéré comme plagié, dans 4 sources.

Liste des sources

Type de source	URL/Nom de fichier	Mémorisé en cache
Source externe/Internet	http://profile.myspace.com/index.cfm?fuseaction=us...	2007-11-17 05:58
Source externe/Internet	http://www.humanite.fr/popup_imprimer.html?id_arti...	2007-11-17 05:58
Source externe/Internet	http://fr.wikipedia.org/wiki/Calvados_(alcool)	2007-11-17 05:58
Source externe/Internet	http://www.humanite.fr/1995-12-08_Articles_-Volail...	2007-11-17 05:58
Source externe/Internet	http://www.forumducommerce.org/news/fullstory.php/...	2007-11-17 05:58

Résultat très surprenant : Urkund détecte bien la source du document, mais partitionne ce document en 9 zones différentes.
Parmi les sources identifiées la première n'a aucun rapport avec le document :

Surprenant aussi que des passages similaires à 100 % soient considérés comme plagiés à 85 %.



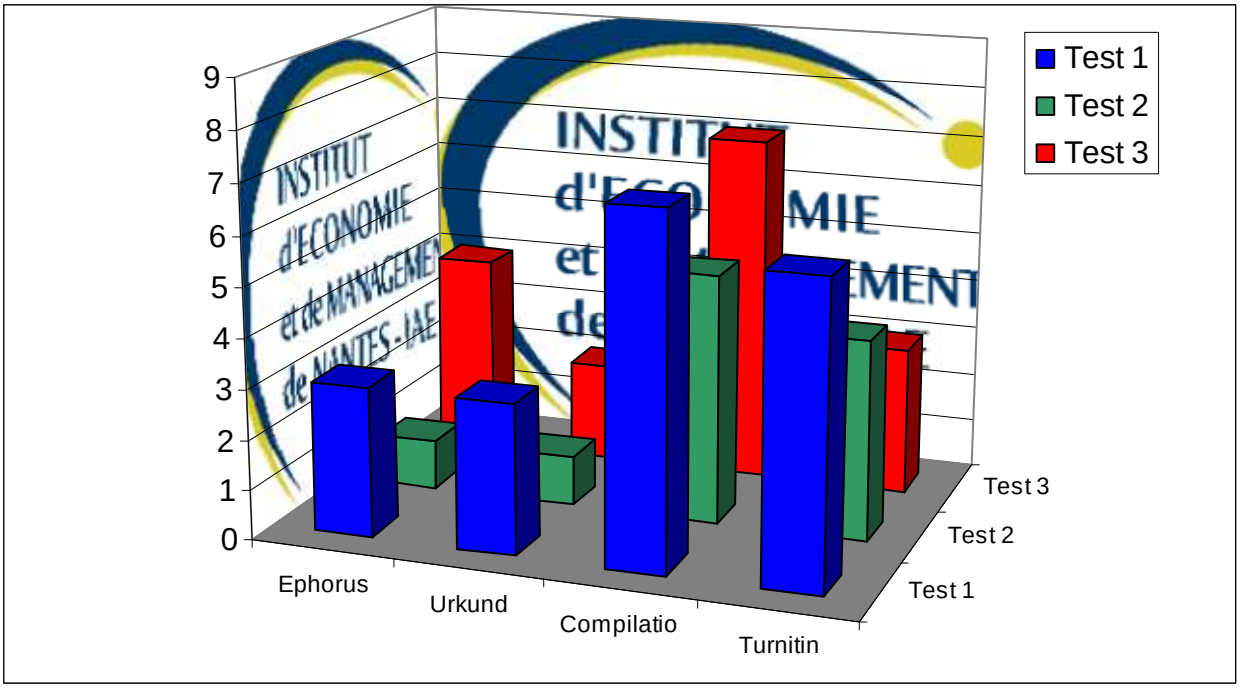
Résultat : 2 zones plagiées détectées sur 9

Document d'origine : Trouvé
Zone 1 : Non trouvée. Blog récent.
Zone 2 : Non trouvée. Site d'information.
Zone 3 : Non trouvée. Article de l'Expansion.
Zone 4 : Non trouvée. Article du Monde
Zone 5 : Non trouvée. Site Oboulo.
Zone 6 : OK.
Zone 7 : Non trouvée. Fichier pdf.
Zone 8 : OK
Zone 9 : Non trouvée. Blog récent.

Conclusion

Quel logiciel choisir ?

	Mots / plagiés	%	Source	**Ephorus 2080 €**		**Urkund 3000 €**		**Compilatio 2390 €**		**Turnitin 2205 €**	
	2476			%	Source	%	Source	%	Source	%	Source
Test 1	338	**14%**	**9**	12%	3	10%	3	22%	7	25%	6
	922										
Test 2	389	**42%**	**6**	2%	1	8%	1	34%	5	58%	4
	2042										
Test 3	546	**27%**	**9**	14%	4	67%	2	30%	7	16%	3



Premier constat : il semble évident que l'ensemble des logiciels gratuits, ainsi que plusieurs logiciels payants, mais peu chers, soient incapables de détecter de façon fiable les plagiats. Par ailleurs ceux-ci nécessitent toujours que l'enseignant importe les documents lui-même, ce qui, à terme deviendra forcément fastidieux.
Restent quatre logiciels : Compilatio, Urkund, Ephorus et Turnitin : c'est-à-dire les deux mieux classés dans l'étude anglaise ainsi que le logiciel français adopté par l'université de Lyon.

<u>En terme de coût :</u>

Si nous nous basons, à titre d'exemple sur un établissement avec 1000 étudiants :

Turnitin : 1435 € pour la licence plus 0.77 € par étudiant soit **2205** € par an. Nombre de soumissions illimitées.

Ephorus : **2080** € pour la licence sans autre droit à verser.

Urkund : environ 3 € par étudiant soit un total de **3000** € pour 1000 étudiants. Ce prix comporte l'ouverture, le suivi et le support de licence. Nombre de soumissions illimitées.

Compilatio : Le coût est en fonction du nombre d'analyse. La zone de stockage coûte 149 € par an. 1000 analyses pour 50 professeurs coûtent **1997 €**. Possibilité d'une licence illimitée à **2398 €.**

<u>En terme de convivialité et simplicité d'emploi :</u>

Sur ce point un critère se révèle très important à l'usage : la possibilité pour le professeur de créer un compte avec des identifiants qu'il remettra à ses étudiants, ceux-ci devant télécharger leur travail avant une date précise. Les quatre logiciels restants ici sont tous pourvus de cette possibilité.
Un autre critère est également partagé par tous : l'effet miroir, c'est-à-dire la possibilité, lorsqu'un plagiat est détecté de mettre en face la source afin de la comparer. Les modes de présentations varient rendant cette fonction plus ou moins rapide, sans pour autant avoir la possibilité de préférer l'une ou l'autre.

Compilatio est très simple à mettre en place et à utiliser. Sans doute le plus intuitif.

Urkund un peu plus compliqué tout en restant très convivial.

Ephorus est très simple d'utilisation, très convivial, d'une grande simplicité et facilité d'utilisation

Turnitin est une 'usine à gaz', comportant beaucoup de possibilités, au détriment de la convivialité. Interface en anglais, sans doute un peu déroutante au début.

<u>En terme d'efficacité :</u>

Il s'agit sans doute du critère le plus important : la capacité d'un logiciel à détecter dans un document supposé original, les parties qui ont été 'empruntées' à d'autres sources d'une façon fiable : tous les plagiats mais rien que les plagiats.
Sur ce point les logiciels ont encore beaucoup de progrès à faire et la recherche de l'exhaustivité amène souvent à la multitude de sources, et donc à l'augmentation des faux plagiats.
Autre point important : la capacité du logiciel à identifier les zones plagiées dans lesquelles les étudiants ont introduits soit des fautes d'orthographes ou des synonymes, dans les but de rendre plus difficile la détection des plagiats. Sur ce point, peu de logiciels sont capables de contourner cette difficulté.
Les algorithmes de vérification étant différent, les résultats le sont forcément.

Compilatio : Compilation compare le document analysé avec l'ensemble des sources disponibles sur Internet, les documents analysés par le passé, et les documents dans les bases documentaires auxquelles l'établissement est abonné.
Sélection aléatoire de suites de 7 mots environ, recherche sur Internet. Si cette suite est trouvée, approfondissement de la source par rapport au document (en théorie). Génère une multitude de sources (325 pour le premier test) dont beaucoup sont redondantes (ce qui

permet également de vérifier qu'Internet est une vaste usine à plagiat) mais surtout dont beaucoup n'ont rien à voir avec le plagiat mais sont le fait de simples coïncidences.
Ceci étant c'est le logiciel qui nous donne les meilleurs résultats puisque 7 des 9 zones plagiées sont trouvées pour le test 1, et une huitième est trouvée en partie (en fait la totalité de la phrase est dans la source identifiée, mais la procédure Compilatio de recherche de 7 mots successifs n'a pas permis de trouver le reste de la phrase et 6 sur 6 pour le test 2, avec l'incapacité à retrouver la source d'origine.
Cette procédure nous découpe (dans le test 1) la zone 9 en 11 zones plagiées, dont toutes n'ont pas le même nombre de sources (9 en général, allant jusqu'à 21 pour la partie 'la plus forte croissance du PIB en ').
Notre document 1 est considéré comme plagié à 22 % (la réalité étant de 13.65 %), le second à 34 % (en réalité 42 %) et le 3ème à 30% (en réalité 27%). Donc globalement de bons résultats, avec l'inconvénient de générer beaucoup trop de 'faux plagiats' et un manque cruel de lisibilité et la totale incapacité de détecter lez zones plagiées modifiées par les étudiants.
Parmi les points positifs du logiciel :
- les résultats d'analyses presque instantanés,
- l'affinage des résultats a posteriori, en ignorant des sources ou des extraits,
- la collecte des documents par e-mail, par "page Internet de dépose", par chargement via l'interface ou encore en interfaçant le logiciel directement avec le site de l'Université.
- une aide à l'utilisation intégrée au logiciel (infobulles) et des tutoriaux animés en français.
Une nouvelle version sera prête pour décembre 2007. Parmi les nouveautés (voir document joint Compilation_New.pdf pour les nouvelles copies d'écran :

Concernant les aspects techniques, le **nouveau rapport d'analyse** prévu pour fin 2007 devrait :
- ignorer (par défaut) les sources présentant des similitudes "accidentelles"
- encadrer les ensembles de textes et indiquer les principales sources pour chaque ensemble (en incluant les petits "inserts" reformulés)
- proposer des solutions de partage des rapports entre enseignants
- mettre en avant les sources présentant le plus de similitude avec le document
- permettre la comparaison du document et de chacune des sources retrouvées

Résultat global :

Test 1 : 7 zones sur 9.
Test 2 : 5 zones sur 6.
Test 3 : 7 zones sur 9.

Turnitin : Turnitin part du principe que 50% du plagiat provient de documents réalisés par des étudiants : la base de données de recherche est donc constituée de 40 millions de documents d'étudiants (écrits en anglais !) auxquelles s'ajoutent 1.2 milliards de pages web stockées et actualisées régulièrement, 10 000 journaux et magazines (souvent en anglais) et des milliers de livres, incluant la littérature classique.
La procédure d'analyse est plus subtile (et secrète), donnant des résultats plus lisibles, et généralement assez performants : notre document 1 comporte, pour Turnitin 30 % de similarité et 38 sources sont identifiées. 6 des 9 zones plagiées sont retrouvées. Il est toutefois très surprenant que la zone 9 ne soit pas identifiée : il s'agit d'une phase de plusieurs lignes que tout moteur de recherche identifie immédiatement.
Pour le document 2, ce logiciel est sans conteste le meilleur : 59 % de plagiat annoncé, 4 des 6 sources identifiées, détection du document de base et contournement des synonymes et fautes

d'orthographe. Deux sources ne sont pas identifiées, sans qu'il soit aisé d'en déterminer la raison.
Pour le troisième test le logiciel Turnitin montre toutes ses limites : ne cherchant pas les informations directement sur Internet mais dans une base constituée à partir d'Internet (donc non exhaustive), beaucoup de sources récentes sont ignorées : ici des informations en provenance de blogs récents ou d'articles de magazines également récents. Seulement 3 plagiats détectés sur 9 !
Globalement : moins de sources proposées, moins de 'faux plagiat'. Excellente lisibilité et mise en miroir des zones plagiées bien pensée, facilitant grandement la recherche. Résultats moins complets que Compilatio, avec l'avantage de ne pas nous noyer dans une masse de sources inutiles. Mais base de données pas suffisamment actualisée pour le marché français (ce qui justifie les excellents résultats de ce logiciel sur les marchés anglais et américains).

Résultat global :

Test 1 : 6 zones sur 9.
Test 2 : 4 zones sur 6.
Test 3 : 3 zones sur 9.

Ephorus : Le document est analysé par le moteur de recherche sur la base de données Internet et la base de données de l'établissement d'enseignement (et éventuellement d'autres bases de données). La procédure d'analyse est plus sophistiquée que Compilatio, avec des résultats pouvant aller du meilleur au moins bons. Sur le premier document les résultats sont corrects, avec une explication plausible sur la raison des zones non retrouvées. Sur le second test, alors que l'auteur met en avant sa capacité à détecter les plagiats avec fautes d'orthographe et synonymes, les résultats sont désastreux ! L'explication donnée par le service technique laisse perplexe : le même test fait par eux en arrangeant le document (augmentation de la taille des zones plagiées) donne de bons résultats.
Pour le troisième document, les résultats sont corrects sans plus (14 % de plagiat pour une réalité de 27%), avec le même défaut que Turnitin : les informations issues de blogs (donc souvent récentes) ou de magazines ou journaux français récents sont souvent ignorées.

Résultat global :

Test 1 : 3 zones sur 9.
Test 2 : 1 zone sur 6.
Test 3 : 4 zones sur 9.

Urkund : la grande déception.
La base de données d'Urkund est importante : Internet public, Internet clos, livres (conventionnels, électroniques, encyclopédies), revues scientifiques et généralistes, ainsi que les travaux d'étudiants stockés dans la base.
Annoncé avec un nouvel algorithme révolutionnaire, lancé il y a 1 mois, les résultats sont loin d'être satisfaisants. Pour le document 1, 3 sources identifiées, dont 1 ne pointe pas vers notre document et 1 n'ayant aucun rapport.
Lors de l'entretien avec le commercial de cette société il paraît évident que ce logiciel est à la recherche de l'excellence sur le critère 'tout le plagiat mais rien que le plagiat'. Mais certainement trop minimaliste pour l'instant.
Le second test a nécessité beaucoup de bonne volonté de ma part pour parvenir à faire analyser le document suite à une erreur technique de leur part : le document proposé deux fois

en quelques minutes indiquait 100 % de plagiat, les deux versions se ‘regardant’ mutuellement. Bonne volonté inutile puisque seulement 1 plagiat détecté sur 6.
Pour le troisième document, la soumission a été plus facile, les résultats sont également meilleurs en étant loin encore de l’acceptable : une source n’ayant rien à voir avec le document, deux zones similaires indiquées à 85% au lieu de 100 % !
Résultats très lisibles, mais manque total d’efficacité sur les documents proposés.
Les techniciens sont alertés : la seule réponse a été de dire que le logiciel est en pleine mutation, une nouvelle version voyant le jour sous semaines, les bugs rencontrés découleraient de ce passage de l’ancienne à la nouvelle version.

Résultat global :

Test 1 : 2 zones sur 9.
Test 2 : 1 zone sur 6.
Test 3 : 2 zones sur 9.

En conclusion : Il convient pour l’instant d’éliminer Urkund tant que les problèmes ne sont pas résolus et Ephorus, les explications sur les faibles résultats n’étant pas satisfaisantes.
Entre Compilation et Turnitin le choix est difficile.
La solution Compilatio sera sans doute plus économique : le commercial n’est pas contre le fait de faire un geste commercial pour contrer ses concurrents. Par contre trop d’information, manque de lisibilité, incapacité à contourner les synonymes et fautes d’orthographe et trop de faux plagiats. Il convient d’attendre la nouvelle version et de faire de nouveaux tests sur ces différents points.
Turnitin, même avec des performances peut être moins évidentes que Compilatio, en terme de densité sur le premier document testé paraît plus à même de donner des résultats utilisables de suite. La quantité est largement compensée par une qualité le mettant en tête des logiciels testés sur la capacité de détection des plagiats modifiés et l’importance des faux plagiats. Par contre la base de données utilisées fait que de nombreux sites récents, blogs, magazines ou journaux sont exclus de la recherche. Idéal pour les Anglais et Américains, sans doute moins pour les Français.