



iThenticate

Logiciel de détection de similitudes de texte

Guide d'interprétation des résultats d'analyse d'un document soumis à iThenticate

par
Marie Lefebvre
Agente de recherche
Décanat des études

Juin 2016

1. Introduction

Le logiciel iThenticate permet de vérifier la présence de similitudes de texte dans un document ou un article que l'on désire soumettre à des fins de publication. Lorsqu'iThenticate vérifie un document, le logiciel compare le contenu du document soumis avec des milliers d'articles disponibles dans ses bases de données. Il est notamment possible de vérifier si une publication est présente dans la base de données d'iThenticate en cliquant sur [ce lien](#).

Lorsqu'iThenticate détecte des similitudes de textes, entre un document soumis et du matériel déjà publié, il surligne chaque passage concerné et il donne le pourcentage de similitude de texte détecté. L'utilisation du logiciel iThenticate est relativement simple donc dans le présent document nous nous concentrerons essentiellement sur l'interprétation des résultats d'analyse d'un document vérifié par iThenticate.

2. Rapport en mode « Text-only »

Les résultats d'analyse du logiciel peuvent être visionnés en mode « Text-only ». Ce mode de visionnement permet d'afficher, les résultats d'analyse en mode texte seulement, c'est-à-dire sans tenir compte de la mise en page initiale. Dans l'exemple ci-dessous, le logiciel a détecté 28% de similarité avec des extraits du texte soumis. Le bouton « Document Viewer » situé en haut de l'écran, donne accès à un menu déroulant qui permet de visionner les résultats sous différentes formes :

- « **Similarity Report / Rapport de similarité** » - Les extraits du texte original sont affichés dans une colonne à gauche. Dans la colonne de droite, on trouve les sources individuelles où du texte similaire a été détecté. Les extraits de texte détectés sont numérotés et ils sont affichés par pourcentage décroissant.

The screenshot displays the iThenticate interface for a document titled "test txt bourse" by Marie Lefebvre. The similarity index is 28%. The interface is in "Similarity Report" mode. On the left, there are two columns of text excerpts from the submitted document. On the right, a list of 8 matches is shown, each with a number, word count, percentage, and source. A red box highlights the "Document Viewer" button in the top left, and a red circle highlights the "Similarity Index" of 28% in the top right. A red arrow points from the list of matches to the corresponding text excerpts on the left.

Match #	Words / %	Source
1	87 words / 14%	Internet from 27-Apr-2010 12:00AM guidesigner.net
2	23 words / 4%	Internet from 22-Jan-2015 12:00AM an.etsmt.ca
3	12 words / 2%	Internet from 13-Jan-2007 12:00AM www.lafredumackelung.com
4	11 words / 2%	Internet from 19-May-2011 12:00AM carteau.uniparis1.fr
5	10 words / 2%	Internet from 24-Mar-2014 12:00AM c3ce.ca
6	9 words / 1%	CrossCheck "Genomics: The Power and the Promise / La génomique - la puissance et la promesse", Genome, 2012.
7	8 words / 1%	Internet psylock.beccalissim.it
8	8 words / 1%	ProQuest "Climate, Water: Gestion des ressources hydriques adaptées aux changements climatiques pour la production optimale d'hydroélectricité. Etude de cas: Barron versant de la zone Manouagan", l'Inquest, 2014.

- « **Content Tracking / Suivi du contenu** » - permet de comparer un extrait du texte soumis avec les différentes sources de textes similaires identifiées par iThenticate.

iThenticate test txt bourse By: Marie Lefebvre Similarity Index 28%

Document Viewer

Mode: Content Tracking

calés. L'objectif secondaire est de déterminer si le choix de paramètres de ces modèles peut être simplifié sans incidence sur la modélisation future. La première phase des travaux consistera en une analyse comparative intra modèle sur des périodes de calage/validation contrastées. Cette analyse sera réalisée sur une base annuelle et saisonnière. À partir des séries hydrométéorologiques existantes, les années historiquement les plus froides et humides (maximisation du ruissellement) et les plus chaudes et sèches (minimisation) seront sélectionnées. La performance de chaque modèle sera établie en fonction de leur habileté à reproduire les débits sur la période de calage, mais surtout en mode validation sur la période contrastée. En seconde phase, les données du Modèle Régional Climatique Canadien (MRCC) seront utilisées en lieu de données observées (1961-2100), aux fins d'élargir la variabilité hydrométéorologique. Ce jeu de données devrait permettre de caractériser de manière optimale la sensibilité structurelle des différents modèles hydrologiques. L'avantage majeur de cette approche réside dans le fait que les données hydrométéorologiques seront parfaitement cohérentes, à l'échelle du bassin, et absentes des erreurs de mesures typiques associées aux données de terrain. Outre une augmentation des connaissances générales dans le domaine de la modélisation hydrométéorologique, ce projet devrait permettre l'apparition de nouveaux modèles plus précis en améliorant leur efficacité. La quantification de la sensibilité des modèles à un climat en évolution permettra de mieux cerner l'incertitude liée aux données historiques, et ainsi mieux prévoir les variations dans le cycle hydrologique en climat futur.

Every day, however, thousands of people are releasing mobile versions of their websites and seeing greatly increased levels of traffic as a result. Unless you optimize your website for mobile phone users soon, you could be left behind by the competition. Creating a mobile-friendly version of your site sounds like a daunting task, but it couldn't be easier to do. Using one of the excellent tools below, you'll be able to build, host and launch your mobile site with absolutely no coding knowledge whatsoever.

There has to be a better way. That's the consensus among American healthcare consumers, who perceive their doctors as behind the times when it comes to their willingness to embrace digital communication technologies like text messaging. Digital technologies, patients note, are widely used in almost every other industry to facilitate transactions and make processes more efficient. Echo est un appareil électronique de

- 87 words / 14% - Internet from 27-Apr-2010 12:00AM [guidesignor.net](#)
This source is #1 in the Similarity report.
- 87 words / 14% - Internet from 13-Jul-2010 12:00AM [www.rsa6.com](#)
This source is completely hidden by one or more sources in the Similarity report.
- 87 words / 14% - Internet from 04-Sep-2011 12:00AM [piemo.amplify.com](#)
This source is completely hidden by one or more sources in the Similarity report.
- 87 words / 14% - Internet from 19-Oct-2010 12:00AM [loislaurentmath.postecous.com](#)
This source is completely hidden by one or more sources in the Similarity report.
- 87 words / 14% - Internet from 02-Sep-2011 12:00AM [teacherlicensingandnewedgewangmenthologes.postecous.com](#)
This source is completely hidden by one or more sources in the Similarity report.
- 87 words / 14% - Internet from 02-May-2012 12:00AM [www.blog-at-lisa.com](#)
This source is completely hidden by one or more sources in the Similarity report.
- 87 words / 14% - Internet from 15-Nov-2014 12:00AM [sylvainstudios.com](#)
This source is completely hidden by one or more sources in the Similarity report.

- « **Summary Report / Rapport sommaire** » - permet d'obtenir la liste de toutes les sources de texte similaires identifiées par iThenticate.

iThenticate test txt bourse By: Marie Lefebvre Similarity Index 28%

Document Viewer

Mode: Summary Report

- 87 words / 14% - Internet from 27-Apr-2010 12:00AM [guidesignor.net](#)
- 23 words / 4% - Internet from 22-Jan-2015 12:00AM [en.etsmtl.ca](#)
- 12 words / 2% - Internet from 13-Jan-2007 12:00AM [www.lettredumarketing.com](#)
- 11 words / 2% - Internet from 19-May-2011 12:00AM [coteq.univ.paris1.fr](#)
- 10 words / 2% - Internet from 24-Mar-2014 12:00AM [cisco.ca](#)

- « **Largest Matches / occurrences élevées** » - permet de voir les sources individuelles qui ont un pourcentage élevé de mot similaires. Quand on clique sur l'hyperlien de la source d'un texte similaire on obtient l'extrait de texte dans son contexte.

Document Viewer

Mode: Largest Matches

modélisation hydrométéorologique, ce projet devrait permettre l'apparition de nouveaux modèles plus précis en améliorant leur efficacité. La quantification de la sensibilité des modèles à un climat en évolution permettra de mieux cerner l'incertitude liée aux données historiques, et ainsi mieux prévoir les variations dans le cycle hydrologique en climat futur.

Every day, however, thousands of people are releasing mobile versions of their websites and seeing greatly increased levels of traffic as a result. Unless you optimize your website for mobile phone users soon, you could be left behind by the competition. Creating a mobile-friendly version of your site sounds like a daunting task, but it couldn't be easier to do. Using one of the excellent tools below, you'll be able to build, host and launch your mobile site with absolutely no coding knowledge whatsoever.

There has to be a better way. That's the consensus among American healthcare consumers, who perceive their doctors as behind the times when it comes to their willingness to embrace digital communication technologies like text messaging. Digital technologies, patients note, are widely used in almost every other industry to facilitate transactions and make processes more efficient. Echo est un appareil électronique de la forme d'un cylindre noir vertical, combinant un haut-parleur à 360 degrés et un assistant personnel à commande vocale, Alexa, capable de répondre à des questions sur les prévisions météo ou les dernières nouvelles, mais aussi de contrôler des objets connectés dans la maison. [1]

Poulin, A., Brisette, F., Leconte, R., Arsenault, R., et Malo, J. (2009)

87 words / 14% - Internet from 27-Apr-2010 12:00AM
galeilongue.net

show in web page

mobile version of your site (free) - search of 2010-04-27 12:00:00 AM - web tools (no download)

Every day, however, thousands of people are releasing mobile versions of their websites and seeing greatly increased levels of traffic as a result. Unless you optimize your website for mobile phone users soon, you could be left behind by the competition. Creating a mobile-friendly version of your site sounds like a daunting task, but it couldn't be easier to do. Using one of the excellent tools below, you'll be able to build, host and launch your mobile site with absolutely no coding knowledge whatsoever.

moblReady (free) You can't build a mobile site using moblReady. What you can do, however, is test your existing site for mobile usability. Enter your URL, and press "Go" to get free reports, detailing how well your site displays on mobiles and any potential problem areas. You're even given information regarding how much people in different parts of the world are charged to view your site. Google Mobile Optimizer (free) Google Mobile Optimizer is the quickest possible means of transforming your website into one suitable for mobile user consumption. Navigate your way to google.com/gsm, enter a site URL, and a lightweight version will appear, without headers, ads or images. It's completely un-customizable and renders a minority of websites useless, but it's a very useful bookmark for your phone. MoFuse (from \$7.95 per month) MoFuse is one of the most popular mobile website builders around, which is unsurprising considering its excellent customizability. Visit the MoFuse website, click "Launch a New Mobile Site", choose a mobile domain and insert your URL. Next, click on "Add Elements" to add dynamic and static content to each page. Click on "Layout" to transfer elements to your live site, dragging and dropping them to create the optimal visual arrangement before, finally, previewing your new site using MoFuse's emulator, which brings up a mobile phone screen on your computer's monitor. For bloggers, MoFuse is completely free and even easier to set up. Mippin (free) Mippin is an absolutely fantastic tool for

3. Rapport en mode « Document Viewer / visionneuse du document »

Le mode « Document Viewer » montre le document soumis dans sa mise en page originale et il affiche les sources individuelles qui contiennent des similitudes de textes identifiées par iThenticate. Pour basculer du mode « Text-only » à « Document Viewer » il suffit de cliquer sur le bouton « Document Viewer » situé en haut de l'écran.

Le mode « Document Viewer » permet d'afficher les résultats en :

- **Mode « Match Overview »** - affichage similaire au mode « Similarity Report » du Text-only (voir point 2)

iThenticate

test txt course
By: Marie Lefebvre

Quotes Included: 28%
Bibliography Excluded

Match Overview

1 Internet 87 words created on 27-Apr-2010 galeilongue.net 14%

2 Internet 23 words created on 22-Jan-2016 4%

3 Internet 12 words created on 13-Jan-2007 www.lrbbedmarketing.com 2%

4 Internet 11 words created on 19-May-2011 cordeau.univ-paris1.fr 2%

5 Internet 10 words created on 24-Mar-2014 c308.ca 2%

6 CrossCheck 8 words 'Genetics: The Power and the Promise / La génétique : le pouvoir et la promesse', Genome, 2012. 1%

7 Internet 8 words project.orca.insm.fr 1%

8 ProQuest 8 words 'Genetics: The Power and the Promise / La génétique : le pouvoir et la promesse', Genome, 2012. 1%

9 ProQuest 8 words 'Genetics: The Power and the Promise / La génétique : le pouvoir et la promesse', Genome, 2012. 1%

La première phase des travaux consistera en une analyse comparative intra modèle sur des périodes de calage/validation contrastées. Cette analyse sera réalisée sur une base annuelle et saisonnière. À partir des séries hydrométéorologiques existantes, les années historiquement les plus froides et humides (maximisation du ruissellement) et les plus chaudes et sèches (minimisation) seront sélectionnées. La performance de chaque modèle sera établie en fonction de leur habilité à reproduire les débits sur la période de calage, mais surtout en mode validation sur la période contrastée.

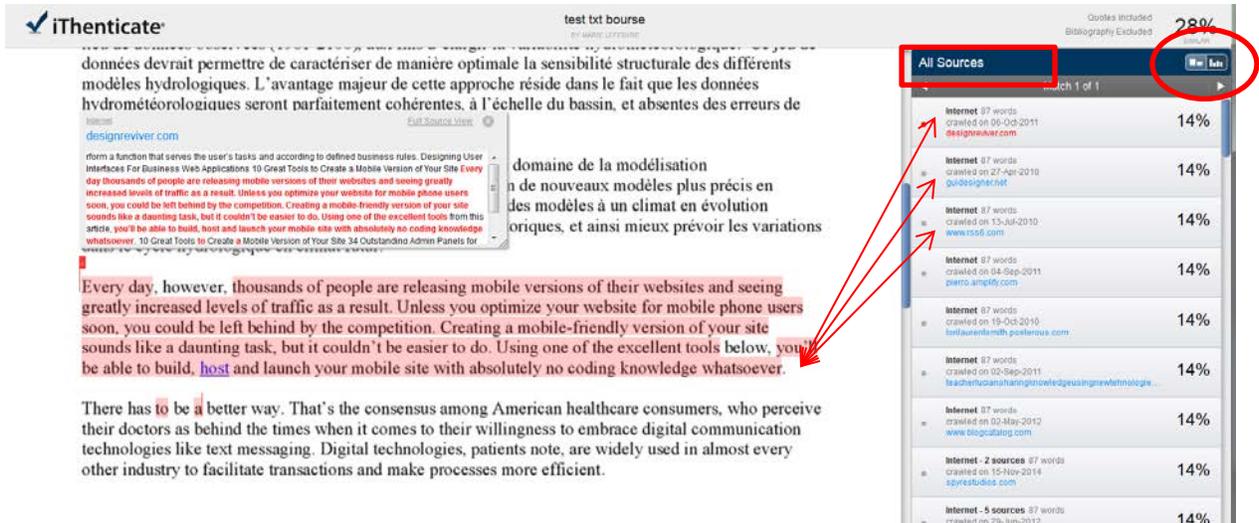
En seconde phase, les données du Modèle Régional Climatique Canadien (MRCC) seront utilisées en lieu de données observées (1961-2100), aux fins d'élargir la variabilité hydrométéorologique. Ce jeu de données devrait permettre de caractériser de manière optimale la sensibilité structurale des différents modèles hydrologiques. L'avantage majeur de cette approche réside dans le fait que les données hydrométéorologiques seront parfaitement cohérentes, à l'échelle du bassin, et absentes des erreurs de mesures typiques associées aux données de terrain.

Outre une augmentation des connaissances générales dans le domaine de la modélisation hydrométéorologique, ce projet devrait permettre l'apparition de nouveaux modèles plus précis en améliorant leur efficacité. La quantification de la sensibilité des modèles à un climat en évolution permettra de mieux cerner l'incertitude liée aux données historiques, et ainsi mieux prévoir les variations dans le cycle hydrologique en climat futur.

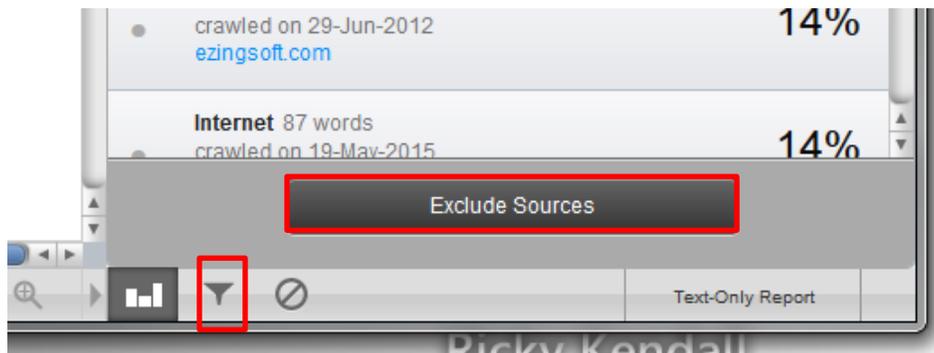
Every day, however, thousands of people are releasing mobile versions of their websites and seeing greatly increased levels of traffic as a result. Unless you optimize your website for mobile phone users soon, you could be left behind by the competition. Creating a mobile-friendly version of your site sounds like a daunting task, but it couldn't be easier to do. Using one of the excellent tools below, you'll be able to build, host and launch your mobile site with absolutely no coding knowledge whatsoever.

There has to be a better way. That's the consensus among American healthcare consumers, who perceive

- Mode « All Sources » - affichage similaire au mode « Content Tracking » du Text-only. (voir point 2)



Il est important de souligner qu'il est possible d'exclure certaines sources de référence en cliquant sur le bouton « Exclude Sources » situé dans la zone au bas de l'écran. L'utilisation d'un filtre permet également d'exclure des citations, des bibliographies ou certains mots. L'icône des filtres est également située au bas de l'écran.



Il est à noter que lorsqu'on exclue une bibliographie, le texte de la bibliographie n'est pas comptabilisée dans le nombre total de mots ce qui donne un pourcentage de similitude plus précis.

4. Calcul du pourcentage de similitude

Le pourcentage de similitude est calculé en fonction du document entier. Le pourcentage de similitude est donc établi en fonction du nombre de mots similaires trouvé dans le document soumis divisé par le nombre total de mots contenus dans ce document.

5. Interprétation du pourcentage global

Il est important de se rappeler que le pourcentage global de similitudes inclut souvent de nombreuses sources (parfois plus de 20) et que chaque source individuelle affiche aussi un pourcentage de similarité. Par exemple, un pourcentage global de 28% peut inclure une source individuelle de 14% de similitude qui comporte un texte substantiel. Tandis que d'autres sources individuelles, par exemple de 1%, peuvent comporter seulement quelques mots de phrases courantes.

Quoted from: Bibliography Excerpt **28%**

test txt bourse
BY NAME, LEFT TO RIGHT

La première phase des travaux consistera en une analyse comparative intra modèle sur des périodes de calage/validation contrastées. Cette analyse sera réalisée sur une base annuelle et saisonnière. À partir des séries hydrométéorologiques existantes, les années historiquement les plus froides et humides (maximisation du ruissellement) et les plus chaudes et sèches (minimisation) seront sélectionnées. La performance de chaque modèle sera établie en fonction de leur habilité à reproduire les débits sur la période de calage, mais surtout en mode validation sur la période contrastée.

En seconde phase, les données du Modèle Régional Climatique Canadien (MRCC) seront utilisées en lieu de données observées (1961-2100), aux fins d'élargir la variabilité hydrométéorologique. Ce jeu de données devrait permettre de caractériser de manière optimale la sensibilité structurale des différents modèles hydrologiques. L'avantage majeur de cette approche réside dans le fait que les données hydrométéorologiques seront parfaitement cohérentes, à l'échelle du bassin, et absentes des erreurs de mesures typiques associées aux données de terrain.

Outre une augmentation des connaissances générales dans le domaine de la modélisation hydrométéorologique, ce projet devrait permettre l'apparition de nouveaux modèles plus précis en améliorant leur efficacité. La quantification de la sensibilité des modèles à un climat en évolution permettra de mieux cerner l'incertitude liée aux données historiques, et ainsi mieux prévoir les variations dans le cycle hydrologique en climat futur.

Every day, however, thousands of people are releasing mobile versions of their websites and seeing greatly increased levels of traffic as a result. Unless you optimize your website for mobile phone users soon, you could be left behind by the competition. Creating a mobile-friendly version of your site sounds like a daunting task, but it couldn't be easier to do. Using one of the excellent tools below, you'll be able to build, host and launch your mobile site with absolutely no coding knowledge whatsoever.

There has to be a better way. That's the consensus among American healthcare consumers, who perceive

Match	Source	Similarity
1	Internet: 67 words created on 27-Apr-2010 guidesignet.net	14%
2	Internet: 23 words created on 10-Jan-2015 en.ete.mta	4%
3	Internet: 12 words created on 13-Jan-2007 www.athromarketing.com	2%
4	Internet: 11 words created on 10-May-2011 cerveau.univ-paris1.fr	2%
5	Internet: 10 words created on 24-Mar-2014 cslce.ca	2%
6	CrossCheck: 9 words "Genomics: The Power and the Promise / La génomique : l'énergie et la promesse", Genome, 2012	1%
7	Internet: 6 words psytech.braca.inseem.fr	1%
8	ProQuest: 8 words Gawlic, D. "Influence du climat et de l'apaisement de la couche organique sur la croissance et la distribution du pe..."	1%
9	ProQuest: 8 words Hajuma, Didier. "Gestion des ressources hydriques adaptées aux changements climatiques pour la production agricole..."	1%

Pour interpréter le pourcentage global obtenu on peut se baser sur les critères suivants :

< 10% Pourcentage faible = Il n'y a pas vraiment de problème

Un taux global de similarité de <10% correspond souvent à quelques mots ou segments de phrases courantes et les sources individuelles indiquent habituellement un taux de similarité de 1 à 3%. De tels résultats peuvent être ignorés.

10% à 50 % Pourcentage modéré = Possibilité de problème (réviser les sources individuelles à pourcentage élevé)

Les documents qui affichent un tel pourcentage global de similarité peuvent contenir des portions de textes copiés littéralement. Il est donc suggéré de vérifier les sources individuelles qui ont +10% de similarité.

>50% Pourcentage élevé = Probabilité de problème sérieux (réviser attentivement)

Un tel pourcentage global requiert une vérification attentive. Il est possible que plusieurs sources individuelles indiquent un haut pourcentage de similarité et il sera nécessaire de les vérifier attentivement.

6. Interprétation du pourcentage des sources individuelles

< 1% - 3 % de similitude = Il n'y a pas vraiment de problème

Ce pourcentage apparaît lorsqu'il y a une séquence de mots ou de courtes phrases qui sont similaires. Généralement, il n'est pas nécessaire de vérifier ces sources individuelles.

4% – 7% de similitude = Prêter attention

Les similitudes peuvent être composées d'une seule phrase ou d'un petit paragraphe. Une source de 6 % n'est pas un problème comme tel, mais quand il y a plusieurs sources individuelles de 6% cela peut valoir la peine de faire une vérification.

8% - 15 % similitude = Vérifier

Lorsqu'une source individuelle indique un tel pourcentage, il peut s'agir d'un ou plusieurs paragraphes qui ont été copiés. Il est donc suggéré de vérifier la source individuelle, car le texte qui apparaît dans le document original n'a peut-être pas été identifié correctement.

15% à 25 % similitude = Vérifier attentivement

Quand une source individuelle a un tel pourcentage, il peut y avoir une quantité importante de texte similaire (ex. : 1 page). Il est donc important de comparer attentivement le document original avec la source individuelle qui a été identifiée.

>25 % de similitude = Attention problème

Une source individuelle qui indique un taux de +25% équivaut à un signal d'alerte! Il s'avère essentiel de comparer attentivement le texte original avec la source individuelle.

Sources de référence :

- Springer Guide How to interpret the results using CrossCheck/iThenticate? 2012.
- CrossCheck Plagiarism Screening : Understanding the Similarity Score
<http://www.ithenticate.com/plagiarism-detection-blog/bid/63534/CrossCheck-Plagiarism-Screening-Understanding-the-Similarity-Score#.Vo2JeVJ8fs4>
- IEEE Cross Check User's Guide, Part 2
<http://www.ieee.org/documents/CrossCheck User Guide S1M119068.pdf>
- IEEE Navigating the CrossCheck/iThenticate Site
<http://www.ieee.org/documents/CrossCheck User Guide.pdf>