

ANÁLISIS DE HERRAMIENTAS PARA LA DETECCIÓN DEL CIBERPLAGIO

Autores: Santos Urbina, Rosa de Ozollo, José Manuel Gallardo, Cristina Martí, Aina Torres, Maria del Mar Torrens.

Palabras clave: deshonestidad académica, plagio, ciberplagio, “copiar y pegar”

Eje temático: Formación para el uso de TIC

Resumen

A lo largo de sus 13 años de experiencia, Campus Extens – UIB Virtual, servicio de e-learning de la Universitat de les Illes Balears, ha ido recibiendo progresivamente más demandas del profesorado relacionadas con problemas derivados de la proliferación de trabajos presentados por el alumnado copiados parcial o totalmente, aprovechando el gran potencial de los medios informáticos.

Desde la UIB ya se ha tratado en profundidad la temática (el MEC financió un proyecto I+D –referencia SEJ2006-10413), si bien nuestra preocupación como servicio es la de poder dar respuesta a esas demandas del profesorado, intentando ofrecer una serie de recursos que le permitan abordar el problema, fundamentalmente, desde las perspectivas de la detección y la prevención.

Para ello estamos elaborando un espacio web en el que mostrar de manera organizada una serie de recursos de apoyo, donde se recogerán: documentos de interés, estrategias metodológicas, acceso a herramientas antiplagio.

En este trabajo presentamos el estudio y análisis de diversas herramientas para la detección de ciberplagio, así como una comparativa que permita, a grandes rasgos, identificar rápidamente las cualidades y debilidades de cada una de ellas. No obstante, hemos de advertir, que en la mayoría de casos se trata de herramientas gratuitas, dadas las dificultades para poder evaluar versiones “demo” de las herramientas comerciales.

Abstract

Throughout its 13 years of experience, Campus Extens Virtual UIB, service that support blended and distance learning at the Universitat de les Illes Balears; has been increasingly getting more teachers demands related to problems resulting from the proliferation of in part or whole copied assets presented by students taking advantage of the great potential of information technology.

The UIB has already been discussed in depth the subject (Project ID Reference SEJ2006-10413 funded by the MEC), but our concern as a service is to be able to respond to these demands of teachers, trying to offer a range of resources that enable teachers to solve the problem from the detection and prevention perspective.

For this we are developing a website that shows in an organized way a series of support resources: documents of interest, methodological strategies, and access to anti-plagiarism tools.

We present the study and analysis of diverse tools to detect cyber-plagiarism, and a comparison that helps to quickly identify strengths and weaknesses of each one of them. However, most tools are free ones; we want to note the difficulties found to evaluate commercial demo tools.

1. Introducció

El Servicio *Campus Extens UIB Virtual*, de la Universitat de les Illes Balears, cuenta con más de una década de experiencia apoyando la puesta en práctica de diversas modalidades de enseñanza universitaria que integran el uso de las TIC en situaciones formativas, tanto presenciales, como semipresenciales y a distancia.

Campus Extens pone a disposición de la comunidad universitaria diversos servicios, como el diseño y la elaboración de materiales didácticos para la red, el soporte técnico a los diferentes usuarios (alumnos y profesores), gestión y organización de proyectos tecnológicos, gestión de videoconferencias, gestión de las aulas virtuales, formación, orientación y apoyo técnico-pedagógico. Este sistema, ha supuesto que el profesorado haya tenido que ir formándose e incorporando progresivamente el uso de las TIC y nuevas estrategias didácticas a su práctica docente.

En este contexto, nuestro servicio ha detectado entre este colectivo una creciente preocupación, especialmente acentuada en estos últimos años, por el ciberplagio. El aumento del número de casos ha disparado las alarmas entre los docentes que, en la mayoría de casos no disponen de herramientas y conocimientos específicos para su detección.

Desde nuestro servicio se plantea pues, a inicios del curso 2009-10, la necesidad de profundizar en el tema a través del trabajo en un proyecto específico dedicado al ciberplagio, con el objetivo de poder ofrecer al profesorado de nuestra universidad estrategias, herramientas y sistemas para la prevención y la detección del ciberplagio.

Nuestra **finalidad** primordial es facilitar a la comunidad de profesores el acceso a recursos con los que hacer frente a este problema desde una perspectiva práctica y directa, y sumarnos así, conscientes de la complejidad del mismo, a las iniciativas puestas en marcha por nuestra universidad en el campo de la investigación de este fenómeno.¹

2. Sobre el ciberplagio

De acuerdo con la *definición* que elaboramos en trabajos precedentes (Urbina, 2004; Urbina, Comas y Sureda, 2006), podríamos caracterizar el *ciberplagio* como:

(...) aquellos materiales presentados como fruto del trabajo personal que, en realidad, han sido copiados parcial o totalmente mediante procedimientos informáticos, sin mencionar su autoría original; la procedencia de los mismos es, mayoritariamente, la red Internet si bien, puede tratarse también de copia de otras fuentes, como enciclopedias en CD-ROM, o trabajos de compañeros.

Para dar cuenta de la extensión del fenómeno, y como ejemplo, cabe destacar que en la investigación llevada a cabo en la Universitat de les Illes Balears, mencionada con anterioridad (Sureda y Comas -dirs., 2008), un 76,6% de los estudiantes de nuestro centro manifestaron haber reproducido en algún trabajo fragmentos copiados de documentos hallados en la web, sin citarlos en modo alguno; esto es, haciéndolos pasar por propios.

Sin embargo, cabría aducir que, aunque el porcentaje mencionado es elevado, no siempre que se produce el ciberplagio ocurre de manera intencionada. Tal como expusieramos en su momento (Urbina, Comas y Sureda, 2006), y de acuerdo con diversos autores (Bugeja, 2001; Park, 2003; Bauer, 2004), cabría diferenciar, pues, de manera clara, dos tipos de ciberplagio:

- Plagio accidental. Se produce cuando se desconoce el procedimiento para la adecuada utilización de fuentes documentales en la elaboración de trabajos académicos y de investigación
- Plagio intencionado. Ocurre cuando hay intención clara de engaño, haciendo pasar por propios textos o fragmentos que no lo son.

¹ En el año 2007 se inició en la Universitat de les Illes Balears un proyecto de I+D financiado por el MEC con el título «El ciberplagio entre los estudiantes universitarios». La investigación fue llevada a cabo por un equipo multidisciplinar e interdepartamental, liderado por el grupo de investigación Educació i Ciutadania de la Universitat de les Illes Balears. Pueden consultarse los resultados del estudio en <<http://ciberplagio.es>>

Pero, ¿cómo puede el profesorado darse cuenta de que un trabajo determinado es producto del ciberplagio? Existen para ello procedimientos muy diversos que van, desde hacer uso de la intuición y la experiencia a partir de la lectura detenida, pasando por hacer un uso manual y minucioso de los buscadores web, intentando localizar fragmentos literales, hasta la utilización de herramientas informáticas específicas para su detección.

No obstante, no cabe duda que, independientemente, del procedimiento utilizado, el énfasis debería recaer sobre las medidas educativas que se pueden adoptar para prevenirlo e intentar minimizar su extensión. Sin embargo, ese será el objeto de un próximo trabajo.

3. Planteamiento del estudio

A pesar del hecho de que el ciberplagio es un problema que afecta a diferentes ámbitos y en diferentes dimensiones, es en el académico donde ha despertado tal vez mayor inquietud, dados los datos que acreditan su extensión, tal como se ha mencionado. Los mecanismos y recursos usados para llevarlo a cabo son variados; pero, en contrapartida, hay que tener en cuenta que, del mismo modo que la tecnología facilita el ciberplagio, no es menos cierto que facilita, a su vez, su detección. Ahora bien, a pesar de la antigüedad de la existencia del plagio, el caso del ciberplagio es de relativa reciente aparición, con lo cual topamos con una falta de conocimiento por parte del profesorado (y del alumnado) de los recursos que pueden tener a su alcance para intentar detenerlo, tanto en el campo de la prevención como en el de la detección.

Así, a instancias de las necesidades detectadas entre el profesorado de nuestra universidad, decidimos llevar a cabo una serie de acciones que permitieran incidir en el problema, para intentar ofrecer nuestro apoyo en la medida de lo posible.

De esta manera, nos marcamos diferentes **objetivos** relacionados con el tipo de recursos a ofrecer, tanto relacionados con la prevención como con la detección y siempre con el previo estudio de la literatura existente sobre el tema, cada vez más abundante.

En cuanto a la *prevención* el objetivo es dotar al profesorado de materiales de apoyo, tales como información legislativa o documentos de estrategias metodológicas, para que éste tenga herramientas para prevenir el plagio (evitar el plagio accidental mediante información al alumnado sobre citaciones...; evitar el plagio intencional a través de cambios en la docencia, evaluación...).

En cuanto a la *detección*, tenemos en cuenta dos ámbitos: la detección personal, con el objetivo de facilitar a nuestros profesores información que les permita a través de indicadores la detección de ciberplagio (cambios de estilos, falta de congruencia...). Y, finalmente, y como parte central del análisis presentado en este trabajo, la detección por *software*; en este caso, se trata de ofrecer la descripción y comparación de una serie de herramientas que puedan ser de utilidad para facilitar esa labor.

Partiendo del proyecto descrito nos centramos en intentar determinar la eficacia del *software* accesible y de los servicios que ofrecen para la detección de diversos tipos de plagio en base a asignaciones de texto y comparar el rendimiento de las herramientas de detección. La intención es, pues, obtener información de primera mano para informar y asesorar acerca de los sistemas de detección más adecuados de acuerdo con los requerimientos de nuestro profesorado. En el siguiente apartado se detallan las fases seguidas durante este proceso.

Una vez se cuenta con toda la información que se desea transmitir, es el momento de decidir cómo hacer llegar a nuestros destinatarios el resultado del estudio. Así pues, se está elaborando una página web con la finalidad de divulgar información y recursos de utilidad práctica (tanto en relación con la prevención como con la detección), seleccionados y organizados para facilitar su consulta y uso, de acuerdo con los resultados obtenidos.

4. Procedimiento: selección de herramientas, análisis preliminar, creación de tablas comparativas.

4.1. Selección de herramientas.

En primer lugar, realizamos una búsqueda en la red con la finalidad de localizar y seleccionar aquellas herramientas que pensamos que podían servir a nuestro propósito. Tuvimos que ir descartando, a medida que las probábamos, las que, ya en un primer momento y por varios motivos, no funcionaban, ya fuese porque el enlace no existía o bien porque no respondía a nuestras perspectivas iniciales. A partir de este amplio abanico, pudimos confeccionar un documento (en formato tabla) que contenía las características básicas y fundamentales que, a nuestro entender, eran necesarias para nuestro estudio. Con estos datos, con este análisis inicial, confeccionamos una ficha (siguiendo una plantilla previamente elaborada) para cada herramienta analizada. Consensuamos la exposición de los resultados, mediante estas fichas, y acordamos mantener y seguir estudiando unas y descartar otras. Los criterios que seguimos, a la hora de escoger unas y eliminar otras, fueron que debíamos rechazar las que, por

ejemplo, no tenían una interfaz intuitiva, las que daban resultados erróneos o poco fiables, las que eran complicadas de manejar... Y continuamos el estudio de las que más se adecuaban a nuestros fines: sencillez, rapidez, usabilidad máxima, resultados verosímiles... a partir de una primera tabla comparativa.

4.2. Análisis exhaustivo de las herramientas elegidas tras la primera selección. Elaboración de fichas detalladas de cada herramienta.

Tras un primer análisis de las herramientas y del descarte de las menos útiles se procedió a un segundo nivel, mucho más detallado y preciso. A partir de aquí, realizamos un análisis exhaustivo de las herramientas: testeando formatos permitidos (.doc, .pdf, .ppt, .jpg, etc.), si ofrecía opciones gratuitas y de pago (y las posibilidades de cada opción), con que idiomas trabaja, manuales de ayuda, el tipo de búsqueda que realiza (principalmente tres: revisando el contenido en otras webs, mirando en una base de datos on-line o mirando en base de datos local y carpetas), tiempos en ejecutar las búsquedas, fiabilidad, y modo de presentar los resultados al usuario (muestra porcentajes, gráficos o texto sospechoso destacado).

4.3. Creación de una segunda tabla comparativa más elaborada y completa

Después de analizar todas las herramientas, se procedió a la creación de una segunda tabla comparativa, mucho más detallada que la inicial, poniendo como ítems todos los elementos y opciones de cada una de ellas, para así tener intuitiva y rápidamente una visión general de lo que nos permiten llevar a cabo y poder compararlas entre sí. En este momento, descartamos aquellas que no cumplían las necesidades de nuestros objetivos iniciales.

4.4. Fichas finales de las herramientas seleccionadas.

Tras el segundo descarte, realizamos fichas individuales que habían pasado el segundo filtro y que ya considerábamos que nos permitían confeccionar un informe claro y conciso para poder llegar a las conclusiones necesarias a la hora de recomendar las más eficaces en la detección del ciberplagio.

5. Análisis comparativo

Tal como define el Virtual Academic Integrity Laboratory (VAIL),² una herramienta de detección de ciberplagio es un *software* o un servicio web a disposición de los

² El Virtual Academic Integrity Laboratory, centro de investigación creado por la University of Maryland University College (UMUC) y desarrollado desde el Center for Intellectual Property and Copyright in the

educadores y estudiantes para facilitar la identificación de un trabajo que se sospecha que ha sido realizado por alguien que no es el estudiante que lo ha presentado. Algunas de estas herramientas buscan texto sospechoso dentro de una base de datos controlada de trabajo de los estudiantes, mientras que otras realizan la búsqueda del contenido en la red (VAIL, 2004).

De acuerdo con esta definición y para facilitar la elección de la herramienta de detección de plagio que mejor se adapte a las necesidades del usuario, se ha elaborado una tabla comparativa que aporta una visión general y ofrece información de las herramientas recomendadas. Tras el proceso descrito anteriormente, estas fueron las finalmente seleccionadas: *Ephorus*, *Approbo*, *Turnitin*, *Plagiarism Detect*, *Plagiarism Checker*, *DocCop*, *Docoloc*, *WCOPYfind*, *EVE2*, *Scriptum*, *Jplag*. A su vez decidimos añadir algunos recursos que, si bien no son herramientas elaboradas específicamente para la detección de ciberplagio, sí que pueden ser de ayuda en la prevención de éste. Es el caso de Image Stamper y Numly.

Así pues, la tabla recoge aquellos ítems que consideramos más característicos y significativos de estas herramientas, a la vez que responden a las necesidades de los profesores usuarios de nuestro servicio, como por ejemplo el hecho que la mayoría de ellos son gratuitos, lo cual permite un acceso más sencillo, no siendo necesaria la adhesión institucional así como el cambio de *software* según las necesidades.

La situación es la que debe determinar la elección de la herramienta, que puede variar en función de las necesidades del profesorado y de las características del alumnado: reciclaje de trabajos de otros cursos académicos o uso indebido de recursos *on-line*. Corresponde al docente la decisión de qué herramienta le será más indicada, de acuerdo con las características presentadas en este estudio.

A continuación enumeramos los ítems seleccionados:

1. **Coste de la herramienta:** indica si la herramienta es gratuita o de pago.
2. **Instalación:** determina si las herramientas son instalables en el propio ordenador, o bien si son accesibles a través de la red.
3. **Ámbito de búsqueda:** se refiere a dónde se realizan las búsquedas, puede ser en Internet, en bases de datos propias, o en bases de datos elaboradas con documentos proporcionados por el servicio.

4. **Carga de archivos:** clasifica las herramientas según la modalidad en que se facilitan los archivos a analizar. Pueden ser documentos completos que se cargan bien en la red, bien en el programa; fragmentos de documentos que podemos cortar y pegar o direcciones web.
5. **Formatos de archivo:** clasifica las herramientas según el tipo de archivos que son capaces de analizar (texto, presentaciones, pdf, web, etc.).
6. **Presentación de resultados:** especifica la forma en que se muestran los resultados del análisis realizado: gráficos, porcentajes, enlaces, documento comparativo y combinación de ellos.
7. **Cuentas de usuario:** para acceder a los servicios de algunas de estas herramientas a menudo se requiere una suscripción, que puede ser individual o institucional.

El resultado de la comparativa en la tabla, una vez testeados todos los programas y revisados todos los ítems que ofrecen los criterios de análisis del *software* evaluado, es el siguiente:

- De las herramientas analizadas, cinco de ellas son de pago, seis son gratuitas y una de ellas tiene una versión gratuita y otra de pago.
- En relación con la instalación de la herramienta, podemos observar que la mayoría son accesibles a través de la web, aunque en el caso de Plagiarism Detect se ofrece tanto esta opción como una versión instalable.
- Tal vez uno de los aspectos que los usuarios tienen más en cuenta cuando se procede a la búsqueda y selección de una herramienta para la detección del ciberplagio sea el ámbito en el que se realizan. Respecto a este ítem, podemos observar que la mayoría las efectúan en documentos que se encuentran en Internet y alguna de ellas en bases de datos. Otro caso es el de las herramientas que combinan las dos opciones, que suelen ser de pago, ya que su ámbito de búsqueda es más amplio. En el mismo orden de cosas, algunas de ellas incorporan los documentos analizados a una base de datos propia, como Turnitin y Ephorus. Caso aparte es WCopyfind, cuya ámbito de búsqueda es el propio disco duro o unidades conectadas en red local.
- La mayoría de las herramientas analizan únicamente archivos en formato de texto. Pocas herramientas detectan ciberplagio en el caso de archivos de imagen, presentaciones multimedia, etc.

- En cuanto a la carga de archivos, podemos considerar que si se da una sospecha general de ciberplagio, lo más conveniente es utilizar herramientas que revisan documentos completos; como se puede apreciar en la tabla, es el caso de la mayoría de las aplicaciones analizadas. Otras herramientas, en las que el usuario selecciona un fragmento de un texto para que sea analizado, se recomiendan en casos de sospecha de párrafos concretos.
- En cuanto a la presentación de los resultados, la mayoría de herramientas combina distintas fórmulas; como podemos comprobar en la tabla, los enlaces a las páginas de plagio son ofrecidos por casi todas ellas, junto con los documentos comparativos y los porcentajes.

Veamos el resultado final en forma de tabla:

TABLA COMPARATIVA DE HERRAMIENTAS DE DETECCIÓN Y PREVENCIÓN DE CIBERPLAGIO

HERRAMIENTA	Coste de la herramienta		Instalación		Ámbito de búsqueda			Carga de archivos			Formatos admitidos					Presentación de resultados				Cuentas de usuario	
	Gratuita	De pago	Web	Instalable	Internet	Base de datos	Compara 2 doc. proporcionados	Cadena de caracteres	Documentos	Direcciones	Texto	Pdf	Imágenes	Presentaciones	Web	Gráficos	Porcentaje	Enlaces	Documento comparativo	Individuos	Instituciones
Approbo	X		X		X				X		X	X		X			X	X	X	X	
Ephorus		X	X		X	X	X		X		X	X		X			X	X	X	X	X
³ Turnitin		X	X		X	X		X	X		X	X		X	X		X	X			X
Plagiarism Detect		X	X	X	X			X	X		X							X		X	X
Image Stamper *	X		X		X					X			X							X	
Numly *	X		X																		
DocCop	X		X		X		X	X	X		X	X			X		X	X	X	X	
Docoloc		X	X		X				X	X	X	X		X	X		X	X	X	X	X

³ Los resultados aquí expuestos en relación con este software provienen de la información proporcionada por Turnitin a través de la red, debido a la dificultad encontrada para realizar una prueba real. Hay que tener en cuenta que, si bien es una herramienta web, para integrarse en Moodle (plataforma educativa utilizada en la Universitat de les Illes Balears) debe instalarse un “programa” en el servidor que puede interferir en algunas de sus funciones. Cabe decir que ya finalizado nuestro estudio ha aparecido una nueva versión de la integración con Moodle que según sus creadores elimina muchos de los problemas iniciales, si bien no podemos dar cuenta de ello.

EVE2		X		X	X	X			X		X						X	X			
Plagiarism Checker***	X	X	X		X			X	X		X							X			
Scriptum		X	X		X	X			X		X						X	X			
WCopyfind***	X			X					X		X						X	X			
Jplag	X			X			X		X		X**						X		X	X	

* Recordamos que éstas no son herramientas específicamente creadas para la detección del ciberplagio, sin embargo pueden ser útiles para la prevención del mismo

** Analiza códigos de programación Java, C, C++, Scheme

*** Herramienta que hace una búsqueda en el disco duro local o en unidades de red, para comparar documentos. No realiza búsquedas en Internet.

6. Conclusiones

Después de haber llevado a cabo el estudio de las herramientas para la detección y prevención del ciberplagio, hemos llegado a las siguientes conclusiones:

- Actualmente existe gran cantidad de aplicaciones que tienen como finalidad la detección del ciberplagio, pero, una vez testeadas, hemos observado que el porcentaje de fiabilidad no se ajusta a la realidad. Por ejemplo, suministrado un texto del que tenemos constancia de un porcentaje de plagio del 80%, lo común es que las herramientas analizadas detecten un porcentaje bastante inferior. Esto se acentúa, especialmente, en las herramientas gratuitas.
- En este mismo sentido, hay que destacar que, aunque se ha avanzado mucho en el tema de la detección del ciberplagio, ésta no se puede garantizar al completo. Es decir, no siempre el plagio se podrá detectar de manera segura a través de las herramientas; también se tendrá que recurrir a otros métodos más tradicionales. Un buen ejemplo de lo indicado anteriormente es el estudio que efectuamos con las aplicaciones escogidas. Comprobamos como las herramientas analizadas, al realizar búsquedas de plagio mediante un único documento, el mismo para todas ellas, nos ofrecían unos resultados que eran muy diversos entre si, habiendo unas diferencias entre el 20% y el 40%. De ahí la hipótesis de que es difícil encontrar alguna que indique en toda su extensión y magnitud el plagio en un documento.
- La mayoría de ellas están encaminadas a la detección de texto, ya sea en .txt, .doc., etc. En cambio para imágenes y para código hay pocas disponibles y, sobre todo, que cumplan con su objetivo.
- La mayoría de las soluciones analizadas usan motores de búsqueda, especialmente Google, para detectar coincidencias; y, generalmente, el tiempo de exploración es casi instantáneo.

URL de las herramientas analizadas

Ephorus <<http://www.ephorus.com/>>

Approbo <<http://approbo.citilab.eu/>>

Turnitin <<http://www.turnitin.com/>>

Plagiarism Detect <<http://www.plagiarismdetect.com/>>

Plagiarism Checker <<http://www.dustball.com/cs/plagiarism.checker/>>

DocCop <<http://www.doccop.com/>>

Docoloc <<http://www.docoloc.de/>>

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

EVE2 <<http://www.canexus.com/>>

Scriptum <<http://www.scriptum.ca/>>

Jplag <<https://www.ipd.uni-karlsruhe.de/jplag/>>

Referencias

Bauer, M. (2004). Accidental vs. Intentional Plagiarism. Avoiding Plagiarism in the classroom. University of Hawaii.

<<http://emedia.leeward.hawaii.edu/resources/plagiarism/03Forms.htm>> (fecha de consulta: 05/09/10)

Bugeja, M. "Collegiate copycats". Editor and Publisher, 2001, 134, p. 22.

Park, C. "In Other (People's) Words: plagiarism by university students-literature and lessons." *Assessment & Evaluation in Higher Education*, vol. 28, 5, October 2003. http://www.lancs.ac.uk/staff/gyaccp/caeh_28_5_02lores.pdf (fecha de consulta: 05/09/10)

Sureda, J. y Comas, R. (dirs.), Casero, A., Gili, M., Urbina, S., Oliver, M., Salvà, F. y Mut, B. (2008). El Plagio y otras formas de deshonestidad académica entre el alumnado de las Universitat de les Illes Balears. Resultados generales. Informe de investigación, disponible en formato electrónico en: <<http://ciberplagio.es/attachment.php?key=37>> (fecha de consulta: 05/09/10)

Urbina, S. (2004): "Ciberplagio: "construyendo" trabajos universitarios." *EDUTEC* 2004. <<http://edutec2004.lmi.ub.es/pdf/159.pdf>> (fecha de consulta: 05/09/10)

Urbina, S., Comas, R. y Sureda, J. (2006). El plagio en la universidad. Análisis documental y de las perspectivas del alumnado y profesorado. *EDUTEC* 2006. <<http://www.sre.urv.es/web/edutec/CDedutec/comunicacions.html>> (fecha de consulta: 06/09/10)

VAIL. Center for Intellectual Property. (2004). *Detection Tools and Methods: A VAIL Guide*. University of Maryland University College.

<http://www.umuc.edu/distance/odell/cip/vail/faculty/detection_tools/intro.html> (fecha de consulta: 06/07/10)