



Integración de plataforma LMS y algoritmo de código abierto para detección y prevención de plagio en Educación Superior

Isnardo Reducindo¹; Luis R. Rivera²; Julio Rivera³; Miguel A. Olvera⁴

Recibido: 10 de julio 2017 / Aceptado: 22 de noviembre de 2017

Resumen. El trabajo muestra los resultados de la primera fase de un proyecto de investigación, enfocado a abordar desde una perspectiva el problema del plagio en la era digital, y como parte de los procesos de formación académica, lo cual se persigue mediante la integración de herramientas tecnológicas. Se trata de una aplicación para detección de plagio construida a partir de dos aplicaciones de código abierto, un LMS y un algoritmo adaptivo de detección de plagio. Se presenta también en términos generales la metodología para el desarrollo, integración e implementación de la herramienta referida, así como el análisis de datos estadísticos a partir de una muestra de estudiantes a nivel superior de la Facultad de Ciencias de la Información (FCI) de la Universidad Autónoma de San Luis Potosí (UASLP), México. El análisis inicial arroja datos de interés y modelos probabilísticos como una primera aproximación al problema, los cuales serán analizados desde un enfoque académico, de manera tal que permita identificar el grado de participación de los estudiantes en este tipo de prácticas como parte de su formación universitaria.

Palabras Clave: Plagio académico; Plataformas de aprendizaje; Código abierto; Algoritmo de detección; Ciencias de la Información.

[en] Integration of LMS platform and open source algorithm for detection and prevention of plagiarism in Higher Education

Abstract. The work shows the results of the first phase of a research project, aimed at addressing from a new perspective the problem of plagiarism in the digital era and as part of the processes of academic training, which is pursued through the integration of technological tools. It is an application for plagiarism detection built from two open source applications, a platform for distance education and an adaptive plagiarism detection algorithm. The methodology for the development, integration and implementation of the aforementioned tool is also presented in general terms, as well as the analysis of statistical data from a sample of students at the higher level of the Faculty of Information Sciences (FCI) of the Autonomous University of San Luis Potosí (UASLP), Mexico. The initial

¹ Universidad Autónoma de San Luis Potosí. Facultad de Ciencias de la Información
E-mail: isnardo.reducindo@uaslp.mx

² Universidad Autónoma de San Luis Potosí. Facultad de Ciencias de la Información
E-mail: rrivera@uaslp.mx

³ Universidad Autónoma de San Luis Potosí. Facultad de Ciencias de la Información
E-mail: jrivera@uaslp.mx

⁴ Universidad Autónoma de San Luis Potosí. Facultad de Ciencias de la Información
E-mail: miguel.olvera@uaslp.mx

analysis yields data of interest and probabilistic models as a first approach to the problem, which will be analyzed from an academic focus, in such a way as to identify the degree of participation of students in this type of practice as part of their university education.

Keywords: Academic Plagiarism; Learning platforms; Open Source; Detection algorithm; Information Science.

Sumario. 1. Introducción. 2. Materiales y métodos. 3. Resultados. 4. Conclusiones. 5. Referencias bibliográficas.

Cómo citar: Reducindo, I., Rivera, L. R., Rivera, J., Olvera, M.A. (2017) Integración de plataforma LMS y algoritmo de código abierto para detección y prevención de plagio en Educación Superior, en *Revista General de Información y Documentación* 27 (2), 299-315.

1. Introducción

En el contexto académico el plagio puede ser entendido como la acción de deshonestidad al momento de copiar en lo sustancial obras ajenas presentándolas como propias. Se refieren también al hecho de extraer, imitar, fusilar o piratear, trozos o ideas de manera fraudulenta sobre obras ajenas de tipo literaria, artística o académicas y no reconocer expresamente al autor original (García: 2000, Moliner: 2004, Martínez de Sousa: 2004, Girón: 2008, Ruipérez: 2016) Así pues, nos referimos al plagio académico como mostrar cualquier tipo de información sin dar el crédito correspondiente al autor original y atribuyéndosela quien la presenta como propia.

Los documentos que pueden ser objeto de esta práctica inadecuada son muy diversos, por ejemplo: tesis, artículos científicos, artículos de divulgación, reportes, libros, novelas, poemas, composiciones audiovisuales, proyectos, planos, maquetas de obras arquitectónicas o de ingeniería, programas de ordenador, bases de datos, entre algunas otras. Dentro de las obras académicas y científicas, algunas de las acciones más comunes de plagio son: prácticas académicas deshonestas relativas al desarrollo de exámenes, ciberplagio, plagio de fuentes impresas, falseamiento de las referencias bibliográficas, falseamiento de datos, entre otras (Macías, 2003; Comas: 2011). Otra clasificación de plagio académico puede ser (Rojas: 2012) plagio directo, plagio por el uso inadecuado de la paráfrasis, plagio complejo usando referencia, plagio con comillas sueltas, el parafraseo como plagio, el auto-plagio o fraude de reciclaje. Estas formas de plagio engloban algunas de las prácticas más comunes en el ámbito académico, como son: omitir la autoría, hacer cambios a escritos originales y presentarlos como propios, no respetar las reglas de la paráfrasis, no indicar con exactitud las páginas de la fuente original, ausencia de comillas y/o reproducir textualmente las ideas del autor, entre otras.

Los factores por los que en el ámbito académico se incurre en plagio pueden ser muy diversos, por lo que no existe una estrategia única para erradicar este problema. Por ejemplo, una encuesta realizada por *U.S. News* y el *World Report Poll* (Cyber Plagiarism: 2017) encontró que el 90% de los estudiantes creen que los plagiadores nunca pagan por las consecuencias de sus actos ya que nunca serán detectados, lo que genera un clima de “confianza” al cometer un acto ilícito dando

como resultado que el 80% de los estudiantes admitan haber plagiado por lo menos en una ocasión. Otra razón que se puede considerar es el desconocimiento de técnicas y metodologías que permiten citar y dar el crédito a cada autor. Además de las razones anteriores por las que se incurre en plagio, también es preciso señalar la tensión que se genera en dos cuestiones: por un lado, proteger la propiedad intelectual y por otro, promover la libertad de creación. Sin duda alguna, es un reto para las Instituciones de Educación Superior (IES) en Latinoamérica atender con precisión cada uno de estos dos aspectos esenciales en el mundo académico y científico, sobre todo en épocas recientes, donde varios casos de plagio llevados a la luz han puesto en duda el prestigio de instituciones y académicos/investigadores consolidados (Cruz: 2015; Avilés, 2015; Milenio: 2015; Martínez: 2015; Redacción Excelsior: 2016) es un hecho que dentro del contexto académico el plagio es algo común a pesar de ser una práctica poco ética y mal vista dentro del entorno (Acevedo: 2015) y que puede incluso llevar a consecuencias legales a quien lo practica.

Por lo señalado previamente, se vuelve imperante que las IES desarrollen propuestas desde el contexto académico y que en su papel formador, atiendan la necesidad de fortalecer la enseñanza de la ética y de los métodos adecuados que permitan sensibilizar a docentes y estudiantes sobre la valía de respetar la creación intelectual de obras ajenas y evitar presentarlas como propias (Astudillo: 2008; Rojas: 2010) En este sentido, es conveniente señalar que el tratamiento de este tema desde el ámbito académico debe ser preventivo y no correctivo, por lo que los esfuerzos que se hagan deben estar enfocados a generar en los futuros documentalistas y académicos, una cultura de respeto a la creación intelectual. Al respecto, un grupo de académicos pertenecientes a doce IES de México, presentaron un documento titulado “*Por una cultura académica distinta: propuestas contra el plagio*” (Zambrano, 2015) en el cual enmarcan ocho puntos críticos que deben ser atendidos en las agendas de las instituciones con respecto al problema del plagio, de entre los cuales se puede resaltar:

“... incluir un compromiso a corto plazo por parte de todas las instituciones participantes en el acuerdo para adecuar su normativa interna a fin de establecer instancias y procedimientos disciplinarios para la detección y, en su caso, sanción a los estudiantes de todos los niveles y a los profesores e investigadores que incurran en esta práctica.”

Para que las IES puedan encaminarse a atender la propuesta descrita en el párrafo anterior y establecer las políticas, herramientas y acciones para la detección, prevención y castigo, se debe tener claro el contexto específico para cada caso; ya que no se puede juzgar igual a un investigador consolidado que plagia un artículo, que a un estudiante universitario que plagia la tarea de su compañero. Dado que este trabajo está enmarcado en el proceso formativo del estudiante universitario, es necesario abordar lo referente a los tipos de documentos que son comúnmente objeto de plagio así como algunos de los tipos de plagio que se pueden identificar y los motivos por los cuales los estudiantes incurren en esta práctica.

1.1. Plagio en el contexto académico

Dentro del proceso enseñanza-aprendizaje, existen una variedad de productos académicos que toman como evidencia la elaboración de documentos para evaluar el desarrollo de competencias en los alumnos, dentro de los cuales se pueden destacar los siguientes.

Tareas (ensayos, reportes, definiciones, etc.): este tipo de documento es la primer instancia en la que los estudiantes comienzan a practicar el plagio, ya que su elaboración puede implicar poco tiempo y a partir de los documentos consultados, el estudiante suele limitarse a “copiar y pegar” contenidos.

Proyectos: este documento es un medio de evaluación con un enfoque de aprendizaje basado en problemas bajo situaciones auténticas o simuladas (Díaz: 2006) usualmente, los estudiantes socializan este tipo de trabajos que previamente ya han sido realizados, lo que conlleva a plagiar entre compañeros, documentos elaborados con antelación.

Tesis: disertación escrita que presenta a la universidad el aspirante al título de doctor en una facultad (Real Academia Española. 2017), en este tipo de documento el plagio resulta un acto inadmisibles en las instituciones de educación superior (Castro: 2013) y que puede llevar a repercusiones graves dentro del ámbito legal. A pesar de esto, las tesis suelen ser uno de los documentos donde frecuentemente se puede identificar plagio (Ruipérez: 2016; Bailey: 2014, 2016) ya que el riesgo puede ser superado por el beneficio de la obtención del grado.

Los tipos de plagio se han diversificado gracias a las herramientas tecnológicas, principalmente Internet, que han impactado en la evaluación de la originalidad de las ideas de los estudiantes (Molina: 2011). Entre los tipos de plagio en que los estudiantes universitarios incurrir, se pueden identificar y describir los siguientes.

Alumno – Alumno: esta modalidad se puede definir como un “reciclaje de trabajos”, en el que los estudiantes comparten los documentos y seleccionan el que cumple con los criterios solicitados, y a partir de éste generan nuevas versiones con pocas o nulas modificaciones.

Alumno – Exalumno: en este caso se identifican acciones referentes a la llamada “herencia académica”, la cual se presenta por el intercambio de trabajos de una misma asignatura en diferentes generaciones, donde el profesor no modifica las características ni los requisitos de los productos solicitados (Comas: 2008) lo anterior facilita a los nuevos pupilos la elaboración al tomar como referencia el trabajo de algún exalumno, lo que concluye frecuentemente en un plagio parcial o total del documento.

Alumno – Autor: este ocurre cuando un alumno plagia un texto de alguna obra protegida con derechos de autor, a la cual pudo haber accedido por medio de la biblioteca o por alguna fuente electrónica.

Alumno – Internet: esta modalidad de plagio es la más recurrente en la actualidad y resulta prácticamente imposible de controlar, ya que en muchas ocasiones provee información falsa debido a que ésta no es verificada en ningún momento por quien la comparte. Surgió a partir de la segunda mitad de la década de 1990, cuando los llamados bancos de ensayos en Internet comenzaron a facilitar

el intercambio de trabajos de manera gratuita o mediante pago (Comas: 2008) y en ocasiones el autor del trabajo resulta ser desconocido. Estos bancos y los procesadores de texto permiten al alumno copiar y reelaborar los trabajos fácilmente de acuerdo a sus necesidades.

Las causas y motivos que dan origen al plagio se pueden atribuir a diversos factores, algunos de los cuales se han abordado en la literatura: el obtener mejores calificaciones y mejores resultados académicos, mala gestión del tiempo dedicado al estudio y a la elaboración de trabajos, por la facilidad de acceso a material vía internet, o por desconocimiento de las normas básicas a seguir para la elaboración de un trabajo académico (Moynihan: 2012) Así pues, para este proyecto consideraremos los siguientes factores.

Desconocimiento: el desconocimiento de metodologías para el correcto uso de fuentes de información, genera que muchos estudiantes consideren que el citar autores demostraría que su texto no es totalmente original, restando mérito a su trabajo (según su punto de vista), además de desconocer como citar correctamente. Un trabajo académico de calidad es aquel que se encuentra documentado y fundamentado, además, un correcto uso de citas y referencias bibliográficas demuestran honradez y conocimiento de las fuentes (Fernández: 2006) algunos al desconocer esta situación optan por la vía más sencilla, la de extraer el texto y apropiarse de él, sin conocer realmente la situación en la que incurrir.

El profesor no revisa a detalle: la revisión que el profesor hace de los productos entregados tiene dos vertientes, por un lado, el cerciorarse de que cumpla con los criterios indicados en cuanto a fondo y forma, y por el otro, considerar el aspecto de analizar y verificar las fuentes y autores consultados, así como las citas integradas en el texto. El segundo proceso es el más complicado y lento, ya que el verificar todas las fuentes de información relacionadas a la temática y compararlas con los trabajos de los alumnos, resulta un desgaste innecesario para el profesor. Además, en ocasiones el plagio puede ser de Alumno – Exalumno o Alumno – Internet, lo que incrementa el número de fuentes a comparar. A pesar de saberlo, el estudiante se ve motivado a incurrir en plagio, ya que las probabilidades de ser descubierto son mínimas.

No existen consecuencias: Al no haber consecuencias se provoca en los alumnos la practicar del plagio, ya que el beneficio que obtienen supera el riesgo de la consecuencia al ser descubierto. Además de que sanciones como la amonestación verbal y pública, o hasta la expulsión de la institución (Ochoa: 2012) no son suficientes como medida correctiva frente un problema masivo que debe ser abordado desde una perspectiva formativa.

Bajo este contexto se enmarca el presente trabajo, que surge con la intención de proporcionar una herramienta integral para prevenir el plagio durante el proceso formativo de los estudiantes universitarios.

La propuesta se refiere a la primera fase y considera una aplicación tecnológica de código abierto basada en una plataforma LMS y un algoritmo de detección de plagio.

2. Materiales y métodos

La idea de integrar una plataforma LMS con una herramienta de detección de plagio no es nueva (Domínguez: 2012) por lo que existen muchas aplicaciones disponibles que cuentan con dicha funcionalidad. Sin embargo, a pesar de que existen distintas plataformas de educación bajo licenciamiento GNU/software libre o de código abierto, no sucede lo mismo con las aplicaciones para detección de plagio, las cuales usualmente cobran por licencia para uso de software o por escaneo individual de documentos. El uso de dichas aplicaciones implica un gasto económico extraordinario que la mayoría de las IES públicas en México, y en general en Latinoamérica, no pueden cubrir ya que no está contemplado como parte de su presupuesto.

Parte fundamental de este trabajo se llevó a cabo la integración de dos herramientas de código abierto, a fin de ofrecer un producto derivado que permita a los docentes la detección y prevención del plagio durante la formación de estudiantes, lo que a su vez les obliga a enfatizar sobre el uso correcto de las fuentes de información.

2.1. Plataformas para gestión del aprendizaje

La tecnología ha modificado con el paso del tiempo el proceso enseñanza-aprendizaje, donde se pueden destacar las plataformas para la gestión del aprendizaje o LMS (Learning Management System), las cuales funcionan como (Sánchez: 2007) herramienta de administración y difusión educativa, que da soporte a la enseñanza y comunicación de forma síncrona y asíncrona entre tutores-alumnos asignados por cursos y grupos de usuarios mediante la creación de ambientes virtuales de aprendizaje. Diversas IES en el mundo, han optado por ofrecer sus cursos de manera remota con ayuda de estas plataformas, o apoyarse en éstas durante algún curso presencial para facilitar las tareas docentes y diversificar las actividades de aprendizaje. Existe una gran variedad de plataformas para este propósito, por lo que es necesario comparar dichas aplicaciones para identificar la más adecuada a partir de las necesidades particulares que se pretenden atender en este trabajo.

Para llevar a cabo la comparación, se utilizaron los datos proporcionados por un estudio previo (Clarenc: 2013) donde comparan 19 plataformas LMS a partir de ocho características: interactividad, flexibilidad, escalabilidad, estandarización, usabilidad, funcionalidad, ubicuidad y persuasión para captar usuarios. Utilizando dichos datos se elaboró la gráfica que se muestra en la Figura 1. De acuerdo a dicha gráfica, se puede apreciar que la mejor opción es Chamilo, y que adicionalmente se adecua a las necesidades de este trabajo, ya que es distribuido bajo una licencia de software libre GNU/GPLv3.

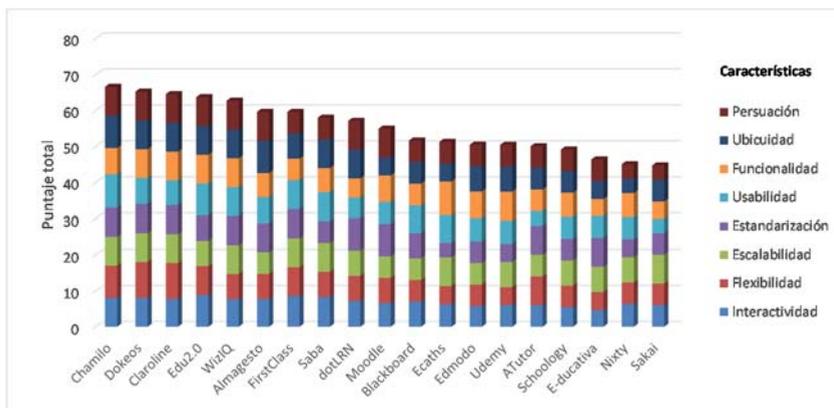


Figura 1. Comparativa de 19 plataformas de gestión del aprendizaje considerando ocho características. Diseño propio a partir de los datos obtenidos del estudio publicado por Clarenc y colaboradores

2.2. Herramientas para detección de plagio

En el contexto académico existen dos alternativas para detectar plagio, una basada en la experiencia y análisis del docente, y otra mediante aplicaciones que encuentran texto plagiado. La primera es la más adecuada, ya que se realiza de manera lógica e identificando de forma concreta las partes en que se incurre en plagio y en qué forma, respetando criterios de citado y parafraseo de acuerdo a normas y estilos bibliográficos establecidos. Sin embargo, cuando el número de trabajos es elevado y además estos son muy extensos, es difícil que el profesor dedique el tiempo suficiente en la revisión, lo que afecta la minuciosidad y calidad de la misma.

Por otro lado, para la segunda opción se dispone de aplicaciones tecnológicas anti-plagio para detectar automáticamente similitudes de texto entre párrafos (análisis intrínseco) o documentos completos (análisis referencial), y se basan en algoritmos desarrollados a través de métodos de extracción/recuperación de información, comparación de caracteres con métodos estadísticos o vectoriales, y análisis de sintaxis, que después son implementados en lenguajes de programación de alto nivel (Introna: 2011) a partir de los resultados presentados por estas aplicaciones, una persona puede entonces determinar si en verdad existe o no plagio.

Existe una gran variedad de aplicaciones que intentan detectar casos de plagio en diversos ámbitos, pero es necesario evaluarlas para elegir la más adecuada según necesidades específicas. Dentro de los criterios a considerar para la selección se encuentran: escalabilidad, precisión en la detección, idioma, requerimientos técnicos, interfaz de usuario, inclusión en algún LMS, licenciamiento en el código fuente y restricciones de uso. Tomando en cuenta estos criterios, se elaboró el cuadro comparativo que se aprecia en la Fig. 2, donde comparan algunas de las herramientas más populares de detección de plagio y que pueden ser utilizadas en el ámbito académico para la revisión de documentos elaborados por estudiantes.

Se analizaron 32 herramientas de detección de plagio, de las cuales dos son desarrolladas desde el ámbito académico y ofrecen el código libre para poder ser modificado y adaptado a las necesidades. La primera de ellas, el *Measure Of Software Similarity* (MOSS) que es para la detección de plagio en el código de software, por lo que se descarta como opción para usarse en el análisis de texto normal; y la segunda, un algoritmo ganador del *Text Alignment Competition on Digital Text Forensic 2014* (PAN 2014), el *Adaptive Algorithm for Plagiarism Detection* (AAPD) (Sánchez-Pérez: 2015) De tal forma que analizando la Figura 2, la única aplicación de análisis de texto que cuenta con el código libre para poder ser modificado y adaptado a nuestras necesidades es el AAPD, además de asegurar la precisión en la detección.

Herramienta	Tipo licencia				Tipo servicio		URL
	Costo	Gratuito Con restricciones	Gratuito Sin Restricciones	Código Abierto	Web	Software	
AntiPlagiarism						✓	http://antiplagiarism.net/
AntiPlagiarist						✓	http://antiplagiarist.softonic.com/
Articlechecker					✓		http://www.articlechecker.com/
Compilatio					✓		https://www.compilatio.net/es/
Copionic					✓		http://copionic.citilab.eu/
Copy check					✓		http://www.copycheck.cc/
Copyscape					✓		http://www.copyscape.com/
Crosscheck					✓		http://www.crossref.org/
Crot software					✓	✓	http://www.crotsoftware.com/
Duplichecker					✓		http://www.duplichecker.com/
Ephorus					✓		http://www.ephorus.com/
Google					✓		https://www.google.com.mx
Grammarly					✓		https://www.grammarly.com/
Heliotext					✓		http://www.heliotext.com/
Ithenticate					✓		http://www.ithenticate.com/
Moss					✓		http://theory.stanford.edu/~aiken/moss/
Paper Rater					✓		https://www.paperrater.com/
Plag Aware					✓		https://www.plagaware.com/
Plag scan					✓		http://www.plagscan.com/es/
Plagiarism checkerX						✓	http://plagiarismcheckerx.com/es/
Plagiarisma					✓	✓	http://plagiarisma.net/es/
Plagium					✓		http://www.plagium.com/
Plagtracker					✓		https://www.plagtracker.com
Quetext					✓		http://www.quetext.com/
Small SEO tools					✓		http://smallseotools.com/plagiarism-checker/
Turnitin					✓		http://turnitin.com/es
Urkund					✓		http://www.urbund.com/en/
Vericite					✓		https://www.vericite.com/
Viper						✓	http://www.scanmyessay.com/
Writecheck					✓		http://en.writecheck.com/
AAPD						✓	http://www.gelbukh.com/plagiarism-detection/PAN-2015

Figura 2. Cuadro comparativo a partir de las características relevantes en herramientas disponibles para detección de plagio. Diseño propio a partir de los datos obtenidos del estudio publicado por Clarenc y colaboradores

2.3. Integración de aplicaciones

Una de las principales razones por las que una gran cantidad de IES en Latinoamérica carecen de alguna herramienta tecnológica que les ayude en la detección de plagio, es por el costo de las licencias. Para este estudio la plataforma de aprendizaje que se utilizó es Chamilo y como herramienta de detección de plagio el AAPD, de acuerdo a los datos presentados en las Figuras 1 y 2 y considerando principalmente que son aplicaciones de código libre, lo que permite modificarlas para integrarlas y adecuarlas a las necesidades específicas de esta propuesta. En la Figura 3 se resumen las características técnicas que se encontraron en las aplicaciones seleccionadas, las cuales permiten visualizar la integración de éstas bajo un entorno de un servidor LAMP (Linux, Apache, MySQL, PHP-Python), y las cuales se fusionaron de acuerdo al diagrama de la Figura 4.

Características	Chamilo	AAPD
<i>Licencia</i>	GNU/GPLv3	GNU/GPL de propósitos académicos
<i>Lenguajes</i>	PHP, JavaScript, HTML	Python
<i>Gestor de Bases de Datos</i>	PostgreSQL, MySQL	N/A
<i>Sistema Operativo</i>	Windows, Linux, MacOS	Windows, Linux, MacOS
<i>Idioma</i>	Multidiomas	Inglés
<i>Comunidad/Aprobación</i>	Nivel Mundial	Ganador PAN 2014

Figura 3. Cuadro que muestra las principales características de la plataforma de aprendizaje Chamilo y el software de detección de plagio AAPD, herramientas integradas para el desarrollo del trabajo. Diseño propio.

La herramienta desarrollada permite comparar todos los documentos que se encuentren almacenados dentro del sistema, brindando al docente la posibilidad de detectar en las tareas y proyectos el plagio Alumno – Alumno, Alumno – Exalumno y Alumno – Internet, este último por medio de documentos que el profesor debe depositar en el sistema. Un ejemplo de cómo está integrado el AAPD a Chamilo se muestra en la Figura 5, y un ejemplo del reporte desarrollado con Python y HTML que muestra los resultados obtenidos con base en el análisis realizado a un documento se observa en la Fig. 6.

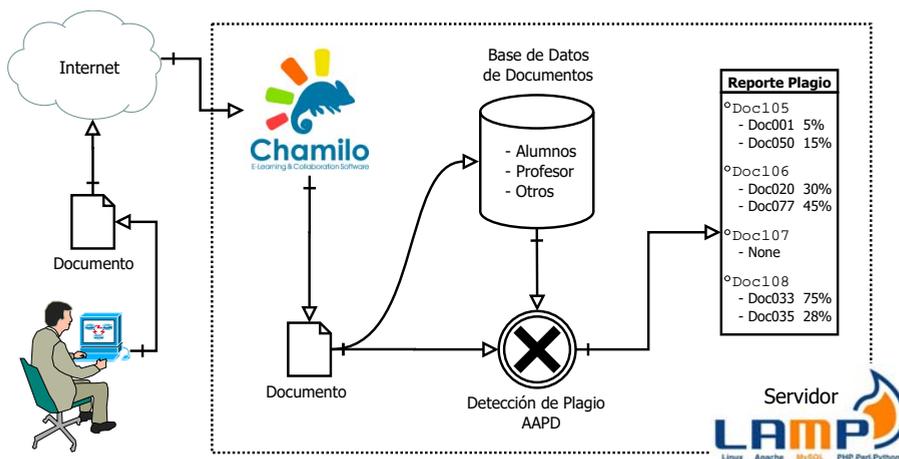


Figura 4. Diagrama de integración de las aplicaciones, así como del flujo del documento en el proceso de revisión de plagio. Diseño propio.

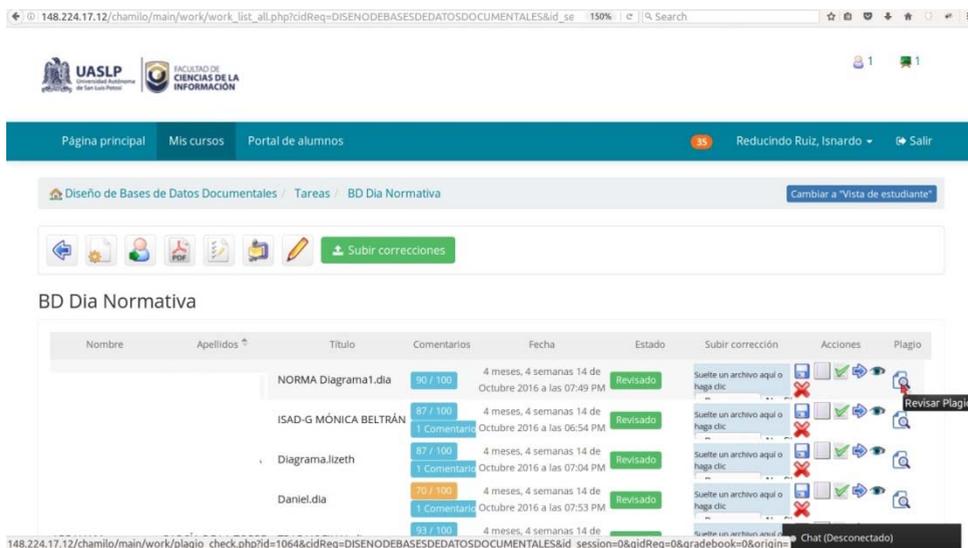


Figura 5. Captura de pantalla sobre la forma en que se observa la integración del algoritmo de plagio de plagio AAPD en la plataforma Chamilo para la revisión de plagio. “Nota. En la imagen –fuente- se eliminaron los nombres de los alumnos” quitar nota...

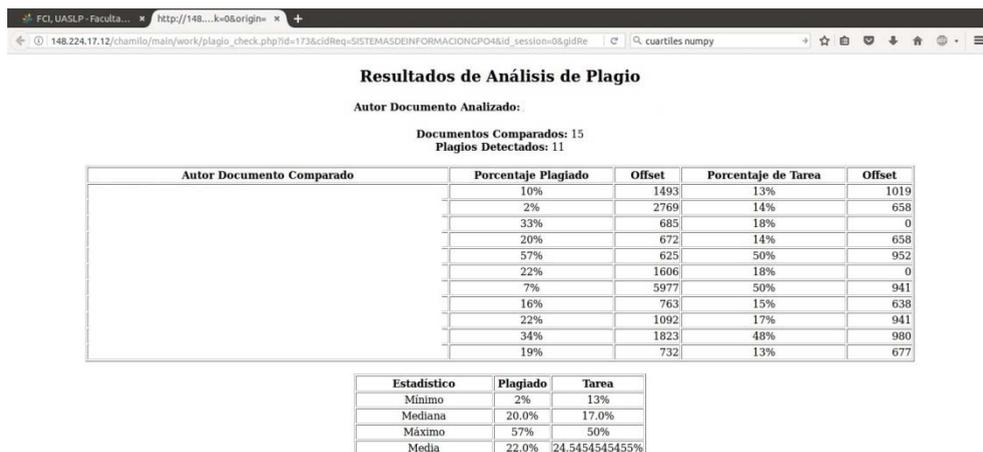


Figura 6. Ejemplo de reporte desarrollado con Python y HTML, obtenido a partir del análisis de un documento mediante la integración de Chamilo y AAPD. “Nota. En la imagen –fuente- se eliminaron los nombres de los alumnos” quitar nota...

2.4. Experimento (muestreo)

Para la recolección de datos necesarios para llevar a cabo el estudio, se determinó trabajar con los grupos de 4 profesores (50% del total) que imparten asignaturas del área de tecnologías de información y comunicación (TIC) en la Facultad de Ciencias de la Información, durante el ciclo escolar 2016-2017, segundo semestre (ene-jun-2017).

Los grupos considerados fueron 6, lo cual representa un 37 % de un total de 16 que integran la comunidad estudiantil en dicha Facultad. Los sujetos que conforman los 6 grupos fueron 113, que equivale al 30 % de un total de 380. Dos de los grupos seleccionados para el muestreo cursaban la asignatura de Diseño de bases de Datos, que se imparte en tercer semestre; dos grupos más conformaron la clase Sistemas de Información, de quinto semestre, y los dos grupos restantes cursaron Proyectos de Automatización en séptimo semestre. Vale la pena señalar que de los 6 grupos estudiados, 3 corresponden al programa de Licenciatura en Gestión de la Información y el resto al programa de Licenciatura en Gestión Documental y Archivística.

Para diseñar el experimento, se solicitó a los estudiantes la integración de un documento en formato de texto plano (txt), en el cual habrían de reportar el significado de 8 conceptos clave para cada una de las asignaturas cursadas. Adicionalmente, se solicitó a los alumnos que redactaran su documento en inglés y lo depositaran en la plataforma *Chamilo*, a través de la herramienta *Tareas*. Por otro lado, cada profesor realizó la búsqueda, selección, análisis y envío de 5 documentos con las definiciones y significados solicitados, mismos que fueron

consultados a través de Google, considerando además, por su relevancia la lista de los primeros 10 resultados.

Los datos estadísticos fueron procesados a través del software *NVivo 11* especializado en el análisis de datos de investigación social. El muestreo realizado representa estadísticamente un margen de error de 6 y un nivel de confianza de 90% respecto a la población de alumnos inscritos en la Facultad durante el periodo del estudio.

3. Resultados

Para los hallazgos del estudio se procedió a la recolección inicial de los datos y procesado de los mismos, con la intención de obtener elementos de corte cuantitativo. A continuación se describen algunos de los resultados alcanzados.

En principio, se procedió a realizar el análisis de los datos recuperados separando por grupos de acuerdo al semestre que cursaban los estudiantes (3º, 5º y 7º). En la Figura 7 se muestra una gráfica de barras agrupada con respecto al semestre, donde se aprecia la frecuencia relativa con respecto a sí los estudiantes incluyeron o no referencias dentro del texto analizado. Al observar la gráfica, es posible identificar que los alumnos de mayor grado son los que respetan la inclusión de referencias dentro del texto.

Además, se procesó la variable de si incluyeron o no referencias en el texto con respecto a la fuente que plagiaron los estudiantes con mayor frecuencia (Internet o un compañero), de lo cual se obtuvo la gráfica de barras que se muestra en la Figura 8. En dicha gráfica se aprecia que de acuerdo a la frecuencia relativa cuando plagian de Internet, es más probable que no incluyan referencias en el texto, y cuando el plagio que se detectó proviene de un compañero, es igual la probabilidad de que incluyan o no referencias.

Otra variable analizada fue el número de fragmentos plagiados en cada texto a partir de lo cual se obtuvo la gráfica de caja y bigotes que se muestra en la Figura 9, donde se agrupa dicha variable con respecto al semestre que cursan los alumnos. Se puede apreciar que el número de fragmentos plagiados por texto disminuye en proporción inversa al semestre que cursa el alumno. Con respecto la cantidad de fragmentos plagiados por texto, se generó un histograma y se complementó con un modelo de distribución Gaussiana para la estimación de la distribución de probabilidad de dicha variable. Lo anterior se puede apreciar en la Figura 10, donde para el modelo se obtuvo una media de $\mu = 4.28$ y desviación estándar de $\sigma = 18.18$.

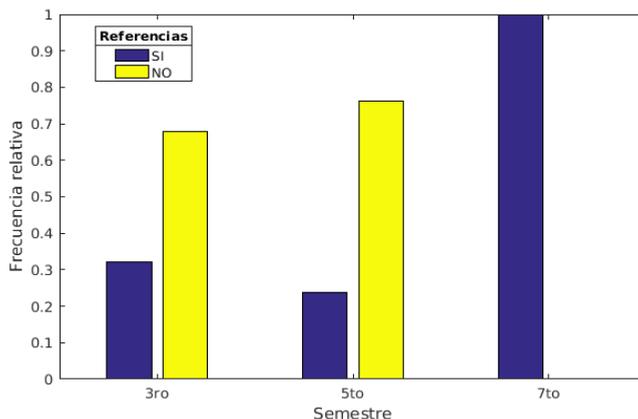


Figura 7. Barras agrupadas de la inclusión o no de referencias en el texto con respecto al semestre que cursan los alumnos de la muestra medida, usando la frecuencia relativa

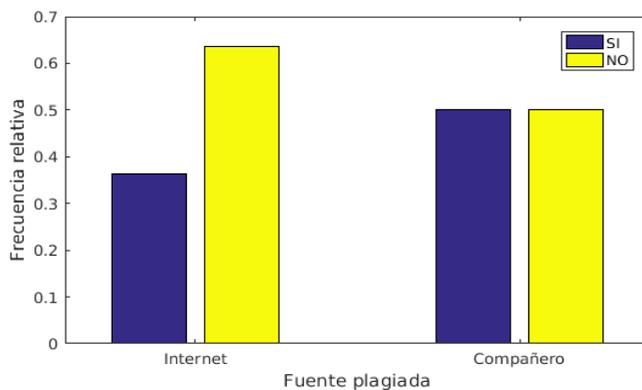


Figura 8. Barras agrupadas de la fuente que plagiaron los alumnos de la muestra con respecto a la inclusión o no de referencias en el texto

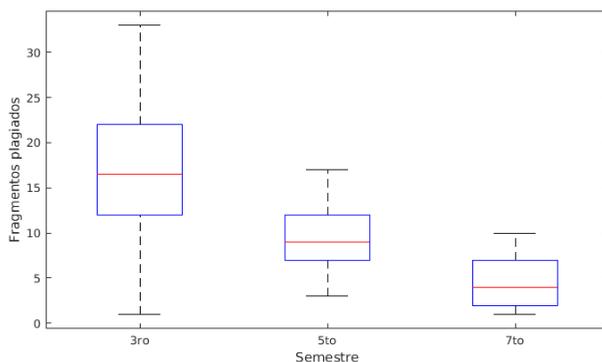


Figura 9. Caja y bigotes del número de fragmentos de texto plagiados por documento de acuerdo al semestre que cursan los sujetos de la muestra

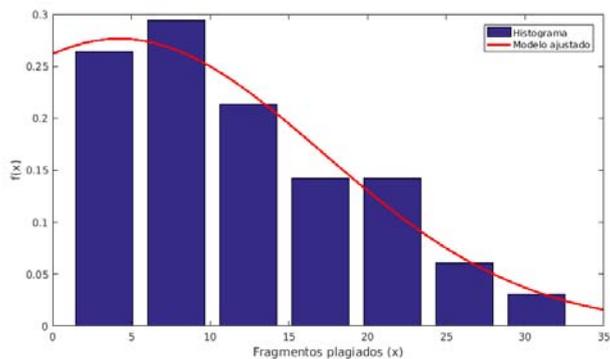


Figura 10. Histograma de número de fragmentos plagiados por documento con curva de ajuste a una distribución de probabilidad Gaussiana

Finalmente se analizó el porcentaje de plagio con respecto a la tarea que representa cada fragmento dentro del texto, para lo cual se elaboró la gráfica de caja y bigotes que se muestra en la Figura 9, donde se agrupó dicho porcentaje de acuerdo al semestre cursado por los alumnos de la muestra. Además se obtuvo de esta misma variable, el histograma y se ajustó un modelo Gaussiano para describir la distribución de probabilidad los cuales se muestran en la Figura 11. La media obtenida que describe la distribución de probabilidad para el porcentaje de plagio por fragmento es de $\mu = 24.68$ con una desviación estándar de $\sigma = 17.66$. De dichas gráficas se puede apreciar que el porcentaje del texto que representan los fragmentos plagiados ronda por el 20%, y que en semestres iniciales es mayor, y caso contrario con niveles superiores pudiendo además comparar con los resultados de la Figura 7, donde se muestra que el 100% de los estudiantes de séptimo semestre incluyeron referencias en su tarea.

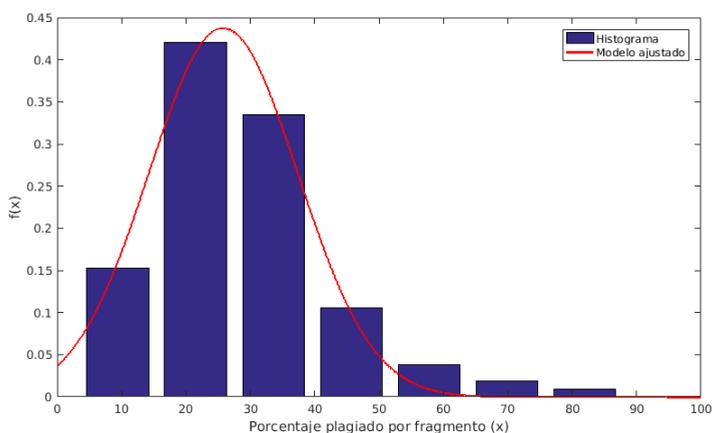


Figura 11. Histograma del porcentaje de plagio por fragmento con curva de ajuste a una distribución de probabilidad Gaussiana

4. Conclusiones

Los resultados alcanzados con el estudio son aún insipientes, sin embargo, dan la pauta para realizar mejoras a la metodología empleada, por ejemplo, ampliar el tamaño de la muestra y diversificar hacia áreas distintas a las TIC, así como extender la investigación a otras disciplinas, buscando con ello lograr elementos más variados, de manera tal que se aproximen a una posible generalización sobre la práctica del plagio entre estudiantes universitarios.

Por otro lado, derivado de los datos estadísticos mostrados, es posible identificar que los estudiantes de semestre más avanzado, habitualmente registran referencias bibliográficas en sus trabajos académicos, a diferencia de aquellos alumnos que cursan niveles inferiores, lo cual permite deducir que a mayor grado de avance, mejores prácticas y hábitos de estudio, por lo menos en cuanto a trabajos escritos se refiere.

Con respecto a las tareas en las que copian y pegan contenidos con mayor frecuencia, se pudo observar que en primer lugar se ubican documentos consultados a través de internet, seguidos de trabajos copiados de los mismos compañeros.

En relación a la cantidad de fragmentos de texto plagiado en cada ejercicio reportado, fue posible identificar, con el apoyo de las herramientas tecnológicas utilizadas, que es mayor entre estudiantes de nivel inferior y disminuye conforme estos avanzan en sus estudios; así, por ejemplo, en alumnos de tercer semestre se detectó que realizan la práctica de copiar y pegar textos sin citar ni incluir referencias, en promedio de 6 a 8 fragmentos (frases, enunciados, párrafos), mientras que los que cursan séptimo semestre solo lo hacen en promedio en 1 o 2 ocasiones.

Resulta necesario destacar la importancia de reforzar las prácticas comunes sobre normas y estilos bibliográficos, a fin de promover entre los estudiantes universitarios el uso correcto de las fuentes de información que les permita desarrollar competencias para facilitar el uso de citas y registro de referencias bibliográficas.

Finalmente, señalar que una de las áreas de mejora detectadas para la integración de la plataforma LMS Chamilo con el algoritmo para la detección de plagio AAPD, es en relación al desarrollo y configuración de librerías complementarias para trabajar en el escaneo de documentos para detectar el plagio no solo en idioma inglés sino en español, así como con archivos distintos a los de texto plano (txt), por ejemplo en formato docx.

5. Referencias bibliográficas

- Acevedo, E. (2015). *La ética y el plagio*. <<https://scratch.mit.edu/projects/71839764/>> [Consulta: 10/03/2017].
- Astudillo, F. (2008) El plagio intelectual. *Revista propiedad intelectual* 5 (8), 242-270.
- Avilés, J. (2015). Investigador de El Colegio de San Luis acusado de plagio. *La Jornada San Luis*, 01-Julio. <<http://lajornadasanluis.com.mx/politica-y-sociedad/investigador-de-el-colegio-de-san-luis-acusado-de-plagio/>> [Consulta: 18/02/2017].

- Bailey, J. (2014) Top Plagiarism Scandals of 2013. *iThenticate Plagiarism Blog* 02 de enero <<http://www.ithenticate.com/plagiarism-detection-blog/top-plagiarism-scandals-2014#.V98f2KJYcxo>> [Consulta: 12/03/2017].
- Bailey, J. (2016) Mexico's President Faces Plagiarism Allegations. *iThenticate Plagiarism Blog* 24 de Agosto <<http://www.ithenticate.com/plagiarism-detection-blog/mexicos-president-faces-plagiarism-allegations#.V98trqJYcxo>> [Consulta: 22/03/2017].
- Castro, S. (2013). *Plagio de tesis y universidad*. México: CETYS.
- Clarenc, C. Ariel; Castro, S. M. (2013) *Analizamos 19 plataformas de eLearning, investigación colaborativa sobre LMS*. <<http://www.congresoellearning.org>> [Consulta: 05/04/2017].
- Comas, R.; Sureda, J. (2008) El intercambio y compra-venta de trabajos académicos a través de Internet. *EDUTEc-e Revista Electrónica de Tecnología Educativa*, 26, 1-16.
- Comas, R.; Sureda, J. (2011) La integridad académica entre el alumnado universitario español. *Revista Estudios pedagógicos*, 37 (1), 207-225.
- Cruz, J. (2015). Denuncian 34 investigadores de Colef plagio académico de dos colegas. *La Jornada Baja California*, 02-Octubre. <<http://jornadabc.mx/tijuana/02-10-2015/denunci-34-investigadores-de-colef-plagio-academico-de-dos-colegas>>[Consulta: 10/02/2017].
- Cyber Plagiarism & Statistics (2017) <<https://www.checkforplagiarism.net/cyber-plagiarism>> [Consulta: 03/01/2017]
- Díaz, F. (2006). *Enseñanza situada: vínculo entre la escuela y la vida*. México: McGraw.
- Domínguez, M. I. (2012) Lucha contra el plagio desde las bibliotecas universitarias. *El profesional de la información*, 21 (5), 498-503.
- Fernández, M. (2006) *La importancia de las referencias bibliográficas y las citas en la elaboración de documentos y trabajos científicos y/o académicos*. <<http://www.bibliotecaminsal.cl/wp/wp-content/uploads/2015/06/Mercedes-Fernandez-Menendez.pdf>> [Consulta: 30/03/2017].
- García, L. (2000) *Diccionario del archivero bibliotecario: terminología de la elaboración, tratamiento y utilización de los materiales propios de los centros documentales*. Madrid: Trea.
- Girón, S. (2008). *Anotaciones sobre el plagio*. Colombia: Universidad Sergio Arboleda.
- Introna, L.; Hayes, N. (2011) On sociomaterial imbrications: What plagiarism detection systems reveal and why it matters. *Information and Organization*, 21 (2), 107-122.
- Macías, H. (2003) *Consejos para evitar plagio y fraudes escolares: y otras plagas similares* <<https://patentesudlap.wordpress.com/2016/11/20/consejos-para-evitar-plagio-y-fraudes-escolares-y-otras-plagas-similares/>> [Consulta: 30/03/2017].
- Martínez de Sousa, J. (2004). *Diccionario de bibliología y ciencias afines*. Madrid: Trea.
- Martínez, G (2015). Nuevo caso de plagio serial en la academia: acusan a investigador de Universidad Michoacana. *El Universal*, 06-Julio.<www.eluniversal.com.mx/articulo/cultura/letra/2015/07/6/nuevo-caso-de-plagioserial-en-la-academia>[Consulta:05/03/2017].
- Milenio Digital (2015). Por plagio, CONACYT expulsa a dos investigadores. El Consejo Nacional de Ciencia y Tecnología expulsó a Rodrigo Núñez Arancibia y a Juan Pascual Gay del Sistema Nacional de Investigadores, por plagiar publicaciones especializadas. *Diario Milenio Digital*, 04-Julio. <http://www.milenio.com/cultura/expulsan-miembros-plagiar-investigaciones_0_566943309.html> [Consulta: 25/02/2017].
- Molina, F.; Velázquez, J. (2011) El fenómeno del plagio en documentos digitales: un análisis de la situación actual en el sistema educacional chileno. *Revista ingeniería de sistemas*, 25 (septiembre) 5-28. 01-Diciembre.
- Moliner, M. (2004). *Diccionario de uso del español*. Madrid: Gredos.

- Moynihan, M. (2012). Plagio: la tentación de fusilar. *Revista Semana*, 18 de agosto <<http://www.semana.com/cultura/articulo/plagio-tentacion-fusilar/263102-3>> [Consulta: 28/02/2017].
- Ochoa, L.; Cueva, A. (2012) El plagio y su relación con los procesos de escritura académica. *Revista Forma y Función*, 27 (2), 95-113.
- Redacción. Acusan de plagio a académico mexicano en España. *Diario Excélsior*, 05 de mayo de 2016. Sección Nacional. <<http://www.excelsior.com.mx/nacional/2016/05/05/1090932>> [Consulta: 25/02/2017].
- Real Academia Española (2017). Diccionario de la lengua española: edición del tricentenario. <<http://dle.rae.es/?id=ZeE19CB>> [Consulta: 15/11/2017].
- Rojas, M. (2012) Plagio en textos académicos. *Revista electrónica Educare*, 16 (2), 55- 66.
- Rojas, M. Á. (2010) Plagio en el ámbito académico. *Revista Colombiana de Anestesiología, Oct/Dic.*, 38 (4).
- Ruipérez, G.; García-Cabrero, J. C. (2016) Plagio e integridad académica en Alemania. *Revista Comunicar*, XXIV, 48 (3), 9-17.
- Sánchez-Pérez, M. A.; Gelbukh, A.; [Et. Al]. (2015) Adaptive algorithm for plagiarism detection: The best-performing approach at PAN 2014 text alignment competition. *Lecture Notes in Computer Science* (9283), 402-413 <https://link.springer.com/chapter/10.1007/978-3-319-24027-5_42> [Consulta: 28/04/2017].
- Sánchez, C. (2007). *Lo que Ud. Nunca se imaginó del e-Learning gestión por procesos*. México: Editorial Porrúa.
- Zambrano, L. (2015) Por una cultura académica distinta: propuestas contra el plagio. *Revista Nexos*. Julio.