

1. DATOS GENERALES DEL PROYECTO

<b>Código:</b>	CIDII-010113
<b>Centro de Investigación:</b>	CENTRO DE INVESTIGACIÓN, DESARROLLO E INNOVACIÓN EN INGENIERÍA
<b>Programa:</b>	Education Assistive Technology
<b>Título del Proyecto:</b>	Desarrollo de un sistema para detección de plagio académico e implementación de una plataforma para investigación de técnicas de detección (SISDEP)
<b>Grupo de Investigación:</b>	Grupo de Investigación en Sistemas Informáticos e Inteligencia Artificial
<b>Area de Conocimiento:</b>	Ciencia y Tecnología
<b>Línea de Investigación:</b>	Sistemas informáticos e inteligencia artificial
<b>Tipo de Investigación:</b>	Aplicada
<b>Campo :</b>	Tecnologías
<b>Investigador Principal :</b>	CRISTIAN FERNANDO TIMBI SISALIMA
<b>Proyectos Vinculados :</b>	
<b>Duración del Proyecto :</b>	12 Meses
<b>Localización del Proyecto :</b>	Cuenca, Ecuador
<b>Fecha de ingreso :</b>	01/10/2013 21:32

## 2. ANTECEDENTES

La Real Academia de la Lengua Española define el término plagiar como ¿Copiar en lo sustancial obras ajenas, dándolas como propias¿, que dentro del ámbito académico extiende a conducta anti ética de estudiantes o profesionales realizada en la redacción de sus tesis, investigaciones, trabajos académicos o escritos, en las que se incluye contenidos intencional o accidentalmente de otros autores sin incluir la cita correspondiente, dándolos como propios. En la actualidad a más de verse como una conducta anti ética desemboca en penalidades académicas y hasta legales constituyéndose en un delito. Delito que atenta contra el reconocimiento y explotación de los derechos de autor.

Las Instituciones de Educación Superior incluyen dentro de sus estatutos o normativas sanciones a actos de este tipo, en las que según la gravedad de la infracción se puede penalizar con una calificación mínima en la valoración de una actividad, una pérdida de un semestre o año académico, llegando hasta dar de baja un título de profesionalización obtenido producto de un trabajo de grado que infrinja en plagio.

La Universidad Politécnica Salesiana en respuesta a la aplicación del artículo 207 de la Ley Orgánica de Educación Superior ha emitido un reglamento para la aplicación de la misma, aprobado por resolución N° 0111-08-2012-12-10, en el que se emiten sanciones por actos dolosos cometidos por los y las estudiantes de la institución.

Específicamente el artículo 7 literal ¿b¿ de este reglamento, considera como una sanción grave a la falta de: ¿Plagiar tesis, investigaciones, trabajos o cuando suplantaren a otros, a fin de obtener notas, certificados y/o títulos, independiente de las acciones penales y civiles contempladas en otras leyes sobre la materia¿; dando en la actualidad una sanción de hasta la pérdida del semestre.

Este comportamiento es un problema a nivel social presente en todos los países del mundo, y día a día se buscan las formas de contrarrestar el mismo, es así como en el internet actualmente se encuentran algunas herramientas que ayudan a analizar plagio académico y así poder validar un documento, a través de la verificación de un plagio textual (uno de los tipos de plagio más comunes), pero dada la evolución de estas conductas también es necesario la evolución de las técnicas de verificación de plagio.

Como se menciono anteriormente el plagio a más de ser una conducta deshonesto, de falta de ética, y tipificada como delito en las legislaciones de algunos países, también constituye un impacto negativo importante en el ámbito académico y profesional, tanto a nivel de la institución que certifica o abaliza el trabajo, como en la pérdida de credibilidad y confianza en el profesional docente o estudiante que comete el delito. De la institución por abalizar un trabajo sin una correcta y completa revisión del trabajo presentado para obtener un título o una calificación para acreditar un conocimiento del tema o aprobación de la materia dentro la carrera estudiantil de una persona; por otro lado el estudiante que crece profesionalmente con un conocimiento vano de la materia y sin contribuir al área de investigación sobre los que se está formando, siendo no merecedor del reconocimiento del trabajo realizado.

## 3. JUSTIFICACIÓN

Considerando los aspectos antes mencionados, se considera importante el estudio y desarrollo de un sistema que facilite el análisis de plagio académico, y que permita una escalabilidad en el mismo, a través de proveer una interfaz que permita desarrollar y probar nuevas técnicas de detección de plagio, así como también algoritmos mejorados de las técnicas existentes y los inicialmente considerados e implementados hasta la finalización del proyecto.

El sistema tiene la finalidad de constituirse en una herramienta de apoyo a la gestión académica-docente y desarrollo investigativo, dirigido para docentes, investigadores y estudiantes como para autores de escritos académicos.

En lo relacionado a la docencia como apoyo en la validación de los trabajos investigativos desarrollados por los estudiantes o en los trabajos de tesis bajo su dirección tanto de pregrado como de posgrado. En lo investigativo y estudiantil apoyo a los actores como autores para hacer una validación de sus trabajos previo a su envío para publicación o valoración por parte del docente, como investigador o estudiante respectivamente.

Desembocando estos aspectos en una mejor producción académica e investigativa.

## 4. OBJETIVOS

### 4.1 Objetivo General

Diseñar e implementar un sistema para la detección de plagio académico e implementación de una plataforma para investigar y probar nuevas técnicas.

### 4.2 Objetivos Especificos

- 1 Estudio del arte de las técnicas de detección de plagio tanto en texto como en imágenes.
- 2 Diseñar e implementar una plataforma genérica para desarrollar y probar nuevas técnicas de detección de plagio en textos.
- 3 Diseñar e implementar un módulo prototipo para detección de plagio de imágenes en documentos académicos.
- 4 Diseñar e implementar un módulo para detección de plagio y análisis de similitud en repositorios de documentos locales.
- 5 Publicación en línea del sistema de detección de plagio académico.

## 5. ESTADO DEL ARTE

### Conceptualización

La Real Academia de la Lengua Española define el término plagio como la *¿Acción y efecto de plagiar¿* [1] con dos enfoques generales: copia y secuestro. Definiendo al término plagiar como:

- \* *¿Copiar en lo sustancial obras ajenas, dándolas como propias¿* [1].
- \* *¿Entre los antiguos romanos, comprar a un hombre libre sabiendo que lo era y retenerlo en servidumbre¿* [1].
- \* *¿Entre los antiguos romanos, utilizar un siervo ajeno como si fuera propio¿* [1].
- \* *¿Secuestrar a alguien para obtener rescate por su libertad¿* [1]

Por un lado se refiere a la copia de obras sean estas de: escritos, fotografías, audio, vídeo y demás valores que atentan contra los derechos del autor de las mismas. Por otro lado se refiere al secuestro de alguien atentando contra sus derechos humanos en relación a la privación de su libertad.

El enfoque que nos compete este proyecto se orienta al relacionado a la copia, copiar sea parcial o completamente una obra (documento escrito) sin consentimiento y reconocimiento del autor de la misma, dándolos como propios; reconocimiento que se logra tan solo con referenciar el real autor por medio de una cita en un formato determinado, esto en el caso de escritos.

Una definición más completa relacionada al plagio en documentos escritos es la que propone la IEEE, donde denota lo siguiente:

*¿Plagiar es reusar las ideas, procesos, resultados o palabras de alguien más sin mencionar explícitamente a la fuente y su autor¿.*

### Tipos de plagio

Dentro del plagio mencionado en adelante llamado plagio académico, existen de manera general 4 tipos fácilmente identificados pero en algunos casos de difícil detección como se menciona más adelante. Los plagios que se referencia son los expuestos por Alberto Barron-Cedeño [2] :

\* Plagio de palabras, tipo de plagio en la que se copia una frase o conocimiento expuesto, sea parcial o totalmente, de un autor sin referenciarlo como fuente, este tipo de plagio puede ser una copia textual o con algunas variaciones aplicando una técnica conocida como parafraseo; parafraseo técnica que se base en re-escribir una misma idea o conocimiento, con palabras diferentes con el mismo significado o alterando su orden. Otra forma también considerada en este tipo de plagio es el plagio o copia traducida, en la que el autor traduce un conocimiento de un autor diferente al de su publicación.

\* Plagio de ideas: tipo de plagio en las que se adopta un conocimiento, ideas, pensamientos o teorías de una persona y se lo da como propio sin referenciar al propietario o autor del mismo.

\* Plagio de fuentes: tipo de plagio en la que se copia la fuentes bibliográficas de un autor sin ni siquiera haberlas leído completamente y también sin citar al autor y el documento de donde las tomo.

\* Plagio de autoría: tipo de plagio en la que un determinado autor suplanta la autoría de otro sobre un documento.

Otros autores mencionan otros tipos de plagio pero que según sus características pueden ser claramente agrupados dentro de los 4 tipos de plagio antes mencionado; por ejemplo Marta Eugenia Rojas Porras catedrática menciona en su artículo [3] tipos de plagio como:

\* Plagio directo o conocido también como plagio textual

\* Plagio por uso inadecuado de paráfrasis, aunque se incluye la fuente original se hacen modificaciones en el texto que no constituye paráfrasis.

\* Plagio con comillas sueltas, en la que aunque se indica la fuente original, se continua exponiendo textos fuera de las comillas.

El incurrir en estos tipos de plagio no siempre es consciente, muchas veces es por desconocimiento por lo que se constituye importante que el escritor se capacite en formatos de citación y de elaboración de paráfrasis.

## Detención de plagio

Conociendo el concepto de plagio y sus diferentes tipos, notamos que la forma de detección de manera manual se hace muy compleja por no decir imposible, pues el conocimiento del revisor debe ser extremadamente amplio y aunque se apoyase de los buscadores web como por ejemplo Google, la tarea tomaría un tiempo alto proporcional a la dimensión del texto en análisis; razones por las cuales surgen los sistemas de detección de plagio, sistemas que apoyados en las tecnologías de información y comunicación automatizan procesos de análisis que surgen de los estudios y desarrollo de técnicas de detección sobre los tipos de plagio detectados.

Los sistemas de detección de plagio de manera general implementan técnicas basadas en experiencia y análisis del docente así como por medio de instrumentos tecnológicos (algoritmos) como lo menciona Marta Eugenia Rojas Porras [3] en la que hace referencia a lo expuesto por autores Comas y Sureda (2007).

Un ejemplo de sistema de detección de plagio básico es el presentado por la revista Ciencia e Investigación Médico Estudiantil Latinoamericana (CIMEL) en la que expone un modelo de detección y tratamiento a casos de plagio, como se muestra en el siguiente gráfico [4] (Axexo 1).

Las técnicas actualmente conocidas y sus modelos computacionales son [2] [5]:

### \* Copia exacta:

Técnica que parte de la premisa que el autor copia un fragmento de texto sin hacer modificación alguna y sin citar la fuente original o exponer el texto entre comillas. En esta técnica consiste en fragmentar un texto en oraciones, frases, o en N-gramas de palabras, para luego proceder a comparar cada una de los fragmentos con una base de datos significativa de documentos relacionados debidamente catalogados o recurrir a buscadores externos, con el fin de encontrar coincidencias exactas.

Los modelos computacionales que mejor responden a esta técnica es los conocidos como huella digital o Hash, en la que a cada fragmento se le somete a un proceso matemático para generar un valor único de correspondencia y con este se compara contra los hash existentes en la base de datos de documentos y sus respectivos conjuntos de fragmentos.

### \*Copia modificada:

Técnica en la que el autor realiza modificaciones sobre un texto antes de incluirlo en su redacción. En esta técnica lo que se busca es encontrar un cierto grado de similitud entre dos textos, en la que tras una operación matemática se calcula la similitud de dos fragmentos estimado un valor de 0 si son completamente distintos o un valor de 1 si son exactamente iguales. Modelos computacionales para esta técnica comúnmente usados son: 1) similitud de coseno, 2) coeficiente de Jaccard.

En esta técnica se consideran técnicas de procesamiento natural de lenguaje, buscando el análisis de esta modificación conocida también como paráfrasis

### \*Copia traducida:

Técnica en la que el autor incluye contenidos de un texto traduciéndolo de un texto originalmente en otra lengua. Una forma de procesar este tipo de plagio es traducir un fragmento a diferentes idiomas y aplicar los modelos antes mencionados.

Estos diferentes tipos de plagio y las conductas seguidas por los autores o escritores son las que se constituyen en la base de la investigación del presente proyecto, pues a más de conocer el tipo de plagio es necesario conocer el proceso de desarrollo del mismo para poder proponer de mejor manera técnicas para su detección, pues a medida de la evolución de las tecnologías que facilitan realizar procesos de detección también se desarrollan nuevas formas para cometer estas faltas.

## 6. METODOLOGÍA

La metodología de trabajo se fundamenta en las metodología de evolutiva de desarrollo de sistemas apoyada en el proceso metodológico de investigación.

Metodología que se constituye de las fases de Análisis, Diseño, Desarrollo, Integración, Pruebas e Implementación, involucrando de manera fundamental la investigación dentro de la fase de análisis (objetivo 1) pues es la que deriva las demás etapas y sirve de punto conexión de las mismas.

Las ventajas de esta metodología y dado la conjunción de subproyectos nos permite seguir secuencias ascendentes o descendentes en las etapas de ejecución del proyecto, así como cumplir etapas o fases en paralelo.(Anexo 2)

## 7. BIBLIOGRAFÍA

- [1] Diccionario de la Real Academia de la Lengua Española, edición 22, (2013), <http://www.rae.es/drae/>
- [2] Alberto Barrón-Cedeño, Marta Vila y Paolo Rosso, ¿Detección automática de plagio: de la copia exacta a la paráfrasis?, 2010
- [3] Marta Eugenia Rojas Porras, ¿Plagio en textos académicos?, Revista Electrónica Educare, Vol. 16, N° 2, [55-66], ISSN: 1409-42-58, mayo-agosto, 2012
- [4] Verónica Rojas-Revoredo, Charles Huamaní, Percy Mayta-Tristán, ¿Plagio en publicaciones científicas en el pregrado: experiencias y recomendaciones?, Revista médica de Chile, (2007), <http://dx.doi.org/10.4067/S0034-98872007000800021>
- [5] Alberto Barrón-Cedeño, ¿Detección automática de plagio en texto?, Masters thesis, Techninal University of Valencia. (2009), <http://repository.dlsi.ua.es/330/>

## 8. RESULTADOS ESPERADOS

- 1 Generar 2 publicaciones científicas donde se presente la propuesta planteada así como los resultados obtenidos a lo largo de las etapas del proyecto.
- 2 Disponer de un sistema de información para realizar análisis de plagio en documentos académicos, el mismo que se encuentre a disposición de los docentes de la universidad así como de otras universidades.
- 3 Disponer de una plataforma para el desarrollo y evaluación de técnicas de detección de plagio, que permitan ir evolucionando el sistema tanto en nuevas técnicas de detección así como en rendimiento y tasa de aciertos,

## 9. TRANSFERENCIA DE TECNOLOGÍA Y/O SOCIALIZACIÓN DE RESULTADOS DE INVESTIGACIÓN

El proyecto de manera general tendrá como resultado dos productos estrechamente relacionados, un primer producto es el sitio web de detección de plagio y un segundo producto es el API (interfaz de programación de aplicaciones) para el desarrollo y pruebas de nuevas técnicas de detención de plagio.

El sitio web se publicará en línea en Internet y se pondrá a disposición de la comunidad universitaria de la universidad para su uso, explotación y validación del sistema, información que contribuirá a la evolución del mismo.

Con respecto a la interfaz de programación de aplicaciones para desarrollar y probar nuevas técnicas de detección de plagio en textos, se generará un documento de especificación debidamente documentado para su uso y desarrollo investigativo.

## 10. IMPACTOS DEL PROYECTO

El proyecto permitirá que tres áreas científicas puedan trabajar de forma conjunta: la psicología, la educación y las ciencias de la computación, generando los siguientes productos:

>> Impactos académicos:

\* Desarrollo de trabajo investigativo en la línea de Sistemas Informáticos/Inteligencia Artificial en el área de detección de plagio.

Mínimo dos tesis investigativas de pregrado de estudiantes de la Carrera de Ingeniería de Sistemas de la Universidad Politécnica Salesiana en el área de detección de plagio.

\* Involucramiento de 3 estudiantes de los últimos ciclos de la Carrera de Ingeniería de Sistemas en el trabajo del proyecto, como labores de pasantías en la industria.

\* Generación de publicaciones académicas.

## 11. INFORMACIÓN DE COFINANCIADORES (en caso de que existieran)

