INSTITUTO POLITÉCNICO NACIONAL
COMUNICADO DE PRENSA
COORDINACIÓN DE COMUNICACIÓN SOCIAL

**México, D.F., a 04 de enero de 2015**

# GANA IPN PRIMER LUGAR EN CERTAMEN INTERNACIONAL DE DETECCIÓN AUTOMÁTICA DE PLAGIO

- **Con un modelo desarrollado en el Centro de Investigación en Computación (CIC) es posible identificar textos plagiados**

- **El certamen se llevó a cabo en la Universidad de Sheffield, Inglaterra**

**C-004**

El alumno de doctorado Miguel Ángel Sánchez Pérez y los investigadores Alexander Gelbukh y Grigori Sidorov del Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN), obtuvieron el primer lugar en la categoría de alineación de textos del *11th Evaluation Lab on Uncovering Plagiarism, Authorship, and Social Software Misuse* (conocido como PAN) celebrado en la Universidad de Sheffield, Inglaterra, por el desarrollo de un modelo de detección de plagio, el cual permite identificar textos producto de la piratería.

El modelo, desarrollado por Sánchez Pérez con la asesoría de Gelbukh y Sidorov para obtener el grado de Maestro en Ciencias de la Computación, superó en el certamen a trabajos desarrollados por competidores de Chile, Estados Unidos, España, Alemania, China y Reino Unido.

Por la aportación tecnológica, con ese mismo modelo, el estudiante politécnico recientemente obtuvo el segundo lugar nacional en el *Concurso de Mejor Tesis en Inteligencia Artificial*, organizado por la Sociedad Mexicana de Inteligencia Artificial (SMIA).

Miguel Ángel Sánchez señaló que descubrir un plagio implica la búsqueda y conocimiento de una amplia cantidad de textos en fuentes originales, por ello científicos de todo el mundo centran sus investigaciones en la generación de modelos para la detección automática de plagio.

Explicó que la localización de fragmentos de texto semejantes entre dos documentos se denomina alineación de textos. Por ejemplo, si el primer párrafo del texto corresponde al tercer párrafo de otro escrito. “Ése es el objetivo del modelo”, apuntó.

Para competir, el modelo debe llevarse a un sistema o software con alto grado de eficiencia, porque se evalúan miles de documentos, se hace un gran número de comparaciones entre textos en busca de fragmentos plagiados. “En el certamen se proporciona a los equipos competidores un corpus (base de datos) aproximado de 5 mil pares de documentos a comparar, los cuales pueden o no contener plagio”, señaló.

Sánchez Pérez comentó que el proceso del concurso consiste en encontrar con el modelo desarrollado los fragmentos similares entre un par de documentos que les proporcionan.

“Para evaluar qué tan bien encontramos un par de fragmentos similares, las medidas usadas son: precisión y exhaustividad. Precisión se refiere a cuántos caracteres del fragmento que detecté realmente fueron plagiados, mientras que exhaustividad se refiere a cuántos, de la cantidad de caracteres que fueron plagiados, detecté. La combinación de esos dos parámetros nos permitió ganar el concurso”, expresó.

Mencionó que la idea de desarrollar el modelo surgió durante una estancia de mes y medio que realizó en la Universidad del Egeo en Grecia con el doctor Efstathios Stamatatos. “Durante mi estancia en Grecia trabajé en aspectos del plagio y me di cuenta que se contaba con recursos para trabajar, había bases de datos, una competencia en la que se podía

participar y eso ayuda mucho a la investigación, porque te permite comparar otros modelos y ver la eficacia de tus sistemas o tus algoritmos", indicó.

Después de que PAN evaluó el modelo y resultó ser mejor que los otros, Sánchez Pérez se dio cuenta que el modelo puede tener alcances importantes. "El sistema podría usarse, por ejemplo, en un administrador de bases de datos de Scopus o de Thomson & Reuters. Cuando se publica un documento el sistema es capaz de decir a qué documentos se parece y solicitar al editor que lo verifique", agregó.

El alumno politécnico señaló que es difícil que un sistema de este tipo tenga una certeza del cien por ciento. "Hace falta la intervención de un humano, pero el sistema puede ayudarle a encontrar textos que quizá no había considerado y con fragmentos específicos para hacerlo más rápido", subrayó.

Sánchez Pérez dijo que además de la detección de plagio, el modelo puede ayudar a la construcción de sitios de contenido colectivo, como Wikipedia, en donde muchas personas escriben artículos, pero se elaboran numerosos contenidos sobre el mismo tema; el modelo podría informar al que escribe si su texto es único o posee similitudes que le permitirían integrarse a otro.

Señaló que a diferencia de otros participantes que no dan a conocer la forma en que obtienen sus resultados, "nosotros tenemos el código abierto en una página del doctor Alexander Gelbukh, por lo que cualquier persona puede acceder y usarlo, sólo tiene que citar el artículo".

**===000===**