

Validando el core del DOCODE

Si bien el año 2010 el equipo liderado por Juan Velásquez se enfrentó a grupos que ya habían competido el 2009, por lo que estaban en desventaja en términos de desarrollo del proyecto, aún así lograron el quinto lugar, lo cual -de paso- sentó un precedente, ya que en esa oportunidad sólo clasificaron dos equipos latinoamericanos (Brasil y Chile).

Este año, con el respaldo de nuevos desarrollos y los algoritmos perfeccionados -además de nuevos computadores-, el equipo DOCODE se propuso ganar la competencia. Y lo logró con creces, ya que casi doblaron en puntaje al equipo que se quedó con el segundo lugar. Un resultado pionero en el área de text mining o lingüística computacional y reconocimiento que valida el *core* del DOCODE.

"Ser los mejores a nivel mundial en detección de plagio nos valida a nivel algorítmico y de investigación", contextualiza Sebastián Ríos, segundo de a bordo de este proyecto.

"Junto con esto, agrega el investigador, también está la componente de negocio que nos exige compararnos con otras herramientas, en particular con Turnitin, software de plagio más vendido en el mundo y utilizado en Estados Unidos, Corea y Japón. Y para hacerlo tenemos que ser mejores y entregar más funcionalidades, lo que trae consigo algunas componentes adicionales: la detección de copia en las redes sociales, la detección de los tópicos en las tareas de los alumnos y la gráfica del DOCODE que es distinta a la de Turnitin, mucho más clara y sencilla de usar y que ha superado con éxito pruebas de usabilidad con profesores en colegios, en Institutos Profesionales y en la misma Universidad".

Detección intrínseca y desafíos futuros

La versión 2011 de esta competencia incluyó un elemento nuevo que, a juicio de los investigadores, es relevante en términos de plagio y que se conoce como 'detección intrínseca de plagio'. Esta herramienta compara un documento consigo mismo y es capaz de determinar variaciones de estilo de escritura, sin necesidad de comparar con posibles fuentes que, debido a Internet, pueden encontrarse en cualquier lugar.

Pese a que los investigadores de Ingeniería Industrial se enfrentaron a grupos con más recursos, los algoritmos del DOCODE fueron superiores. Un trabajo que fue calificado como de *outstanding achievement* por los referies de la competencia quienes, cuenta Velásquez, quedaron asombrados por la distancia en los resultados que lograron respecto del equipo que obtuvo el segundo lugar.

Sobre lo que viene a futuro, los investigadores responsables de este proyecto adelantan que el próximo año volverán a competir por el primer lugar.

A corto plazo, en tanto, planean incorporar más idiomas al software para que éste sea capaz de detectar un párrafo copiado en otro idioma (actualmente el DOCODE opera sólo en inglés o en español).

Junto con esto, también planean incluir la detección de texto oculto que son espacios rellenos con letras en color blanco para que el software de detección de plagio lo lea

como un todo (texto completo) el cual, al no tener sentido, impide que el sistema detecte la copia. Una práctica que se comenzó a pesquisar en el año 2009 en algunos países del mundo, entre ellos Rusia.

"Ya implementamos un primer módulo que permite detectar si un texto viene preparado maliciosamente, aplicación que no ofrece Turnitin y elemento que le agrega robustez al DOCODE", puntualiza Gabriel quien también es alumno del Magíster en Gestión de Operaciones (MGO).

Agrega:

"Si bien trabajamos la problemática del plagio desde el punto de vista de detección, o sea, cuando la copia ya está ocurriendo, existe un ámbito que es la prevención de esta conducta y que tiene que ver con la educación, el código de honor y la labor de los profesores en él. Es muy distinto que un docente pida definiciones de algunas materias a que solicite aplicaciones de ella, ya que éstas últimas son más difíciles de plagiar. Esto va de la mano de la prevención que es un área que nos gustaría desarrollar en un proyecto futuro", analiza Gabriel.

Sebastián complementa:

"La idea es que ya que tenemos una herramienta que detecta plagio podamos capacitar a los profesores en su uso y los orientemos sobre cómo generar tareas para que los alumnos aprendan de mejor manera".

"Se trata de dar una solución completa al tema del plagio, que va más allá de la detección", concluye Gabriel.

Competencia PAN

La Competencia PAN (Uncovering Plagiarism, Authorship, and Social Software Misuse) se creó en el año 2007 y el 2008 organizó su primera versión, en esa oportunidad a nivel privado. En el 2009 se abrió al público en general, año a partir del cual cualquier equipo del mundo puede participar en ella.

"Esta competencia otorga un *framework* a nivel mundial y eso es clave. Por otra parte, los participantes son evaluados por un grupo externo en un proceso muy transparente y transversal", destaca Gabriel Oberreuter, Ingeniero Civil Industrial y 'cerebro' de los algoritmos del DOCODE.