

Desarrollan un nuevo método para la detección automática de plagios de texto

Expertos del Laboratorio de Ingeniería en Lenguaje Natural (Lab NLE), integrado en el Grupo de Ingeniería del Lenguaje Natural y Reconocimiento de Formas (ELiRF) de la Universidad Politécnica de Valencia, han desarrollado un nuevo método para la detección automática de textos plagiados y, en concreto, de plagio traducido. El trabajo, publicado recientemente en *Languages Resources and Evaluation Journal*, se enmarca dentro del proyecto Text-Enterprise 2.0, financiado por el Ministerio de Ciencia e Innovación.

UPV

5/4/2011 10:52 CEST



Paolo Rosso y Alberto Barrón-Cedeño, en el Laboratorio de Ingeniería en Lenguaje Natural de la UPV. Foto: IPV.

“En nuestros días, el problema del plagio, en particular el de texto, se ha incrementado debido al fácil acceso a grandes fuentes de información a través de medios electrónicos. Desafortunadamente, su detección es prácticamente imposible de forma manual. Por ello, es importante desarrollar mecanismos automatizados que permitan realizar la tarea de detección de plagio y así combatir la creciente tentación de plagiar desde la

Web”, apunta Paolo Rosso, investigador del Lab NLE de la Politécnica de Valencia.

Junto a Alberto Barrón-Cedeño, estudiante de doctorado de la UPV, Paolo Rosso trabaja en la actualidad en el desarrollo de tecnología y aplicaciones para la detección automática de plagio y concretamente de plagio traducido, por ejemplo cuando la fuente está escrita en inglés y el plagio está en castellano o valenciano.

“El plagio traducido se hace más frecuente cuando la información que se busca no está disponible en la Web en el idioma materno, por ejemplo el valenciano. Si la detección de plagio en una misma lengua es de por sí complicada, ya que una persona puede modificar el texto original, cuando esta modificación implica un cambio de lengua la dificultad es aún mayor. Pocos son los métodos que han sido desarrollados para abordar este tipo de plagio”, explica el profesor Rosso.

Los investigadores de la UPV han desarrollado uno de estos métodos, que se basa concretamente en modelos de traducción estadística. Estos modelos “aprenden” cuáles son las potenciales traducciones de un texto de un idioma a otro a partir de grandes colecciones de documentos.

“Con lo aprendido es posible calcular la similitud entre textos escritos en distintos idiomas y, si esta similitud es muy alta, se puede sospechar de la existencia de un caso de plagio traducido”, apunta Alberto Barrón-Cedeño.

A diferencia de otros modelos existentes, el método desarrollado desde el Laboratorio del NLE de la Politécnica de Valencia tiene la ventaja de ser robusto cuando se enfrenta a lenguas poco relacionadas sintácticamente (por ejemplo, castellano y euskera).

“El objetivo es proporcionar la evidencia necesaria para que un experto tenga las mejores condiciones posibles para tomar una decisión final respecto a si ha habido o no plagio”, añade Paolo Rosso.

Competición Internacional

En el marco de esta línea de investigación, el Lab NLE de la Universidad

Politécnica de Valencia ha organizado, conjuntamente con la Bauhaus Universitat Weimar (Alemania), y por tercer año consecutivo, una competición internacional sobre detección automática de plagio (<http://pan.webis.de/>).

Patrocinada por Yahoo! Research, esta competición tendrá lugar en Amsterdam el próximo mes de septiembre, dentro del foro europeo de evaluación del CLEF (<http://clef2011.org/index.php?page=pages/labs.html>).

Derechos: **Creative Commons**

TAGS

TRADUCIDO | PLAGIO | TEXTO | UPV | LENGUAJE | LENGUA |

Creative Commons 4.0

Puedes copiar, difundir y transformar los contenidos de SINC. [Lee las condiciones de nuestra licencia](#)