

# Minería de datos en textos médicos para identificar pacientes con similares características

**PERSPECTIVAS.** Los sistemas electrónicos hacen posible que exista una gran cantidad y diversidad de textos médicos digitales (informes de alta, informes radiológicos, etc.) que contienen información de medicamentos, enfermedades y tratamientos.

Son textos que carecen de un formato estructurado o semiestructurado ya que están escritos en forma de narrativa y contienen terminología especializada, abreviaturas, acrónimos, negaciones, etc. Estas particularidades dificultan la identificación de entidades médicas y, más aún, hacer un análisis manual de relaciones que pueden existir entre la gran cantidad de documentos que se generan en el área de la atención y cuidado de la salud.

Ruth María Reátegui Rojas se ha propuesto indagar sobre métodos y técnicas de minería de datos que permitan encontrar similitudes en los textos clínicos y lidera un proyecto de investigación que extrae información de diagnósticos médicos, como enfermedades y síndromes, desde documentos de varios pacientes distantes entre sí, pero que sufren una enfermedad específica.

“Luego —señala— se aplican métodos y técnicas de minería de texto para determinar similitud entre documentos que, por ejemplo, identifica a grupos de pacientes con características similares”.

Profesora de la Sección de Inteligencia Artificial en el Departamento de Ciencias de la Computación y Electrónica de la UTP, Ruth María Reátegui reconoce que “analizar documentos médicos tiene muchas implicaciones éticas puesto que estamos trabajando con información de seres humanos” y señala que, por eso mismo, “es complejo obtener este tipo de documentos directamente de instituciones de salud. Por tanto, para este proyecto, se

obtuvieron los documentos médicos de organizaciones que han puesto a disposición este tipo de información únicamente con fines investigativos”.

No obstante, hasta el momento ya se ha realizado la extracción automática de enfermedades reseñadas en cada uno de los documentos médicos analizados y, aplicando técnicas de minería de texto y *machine learning* o aprendizaje automático de máquinas, se han podido identificar grupos de pacientes con enfermedades similares.

“Aún queda mucho por investigar”, afirma la profesora Reátegui, “y siempre hay que mejorar los resultados considerando nuevas características de los pacientes, como por ejemplo los medicamentos, los tratamientos, etc”.

“También consideramos que es necesario explorar nuevas técnicas y aplicar nuevos algoritmos, pero creemos que los resultados que se obtengan en grupos de pacientes con enfermedades similares serán de utilidad y permitirán, por ejemplo, elaborar una guía para los médicos donde se recomiende tratamientos a nuevos pacientes que tengan características similares a los grupos encontrados, porque la información obtenida de esos grupos puede ayudar a predecir la salud”, añade.

El equipo de investigación, en el que colabora personal de la *École de Technologie Supérieure* (Canadá), no descarta que en un futuro sus aportaciones informáticas permitan llegar a crear un sistema que, a través del análisis de los datos de cada informe médico, identifique pacientes de un grupo específico con características lo más similares posible, en el que incluso se pueda disponer de herramientas que asistan a los doctores en la selección de tratamientos.