

## INGENIEROS DESARROLLAN RECONOCIMIENTO AUTOMÁTICO DE MOTIVACIONES TRAS LAS PREGUNTAS DE YAHOO! ANSWERS



Como base se usó Yahoo! Answers, que cuenta con más de 400 millones de preguntas.

Los servicios de preguntas y respuestas comunitarias (community Question Answering, cQA) son ampliamente utilizados por usuarios de internet, ya que les permite fácilmente realizar cualquier tipo de pregunta y obtener luego múltiples respuestas de otros sujetos de la comunidad. Sus millones de miembros, esparcidos por el mundo, ingresan continuamente sus preguntas y respuestas en estos sitios que gradualmente se convierten en repositorios a gran escala de información reutilizable.

Existiendo escasa indagación en la literatura al respecto, investigadores de la Universidad Andrés Bello se propusieron comprender y, particularmente, detectar automáticamente lo que motiva a las personas a hacer preguntas a desconocidos como una manera de obtener información pertinente a sus necesidades, en lugar de recurrir a un motor de búsqueda. Para este proyecto se utilizó como base el foro Yahoo! Answers, que cuenta con más de 400 millones de preguntas en su repositorio. Los resultados de este trabajo permitirían que, a través de etiquetas que proporcionen información adicional, los usuarios respondan conociendo la situación contextual de la necesidad y cómo se evaluará la calidad de la información, y desarrollar sistemas para dirigir preguntas a expertos sobre la base de las motivaciones del solicitante, entre otras aplicaciones.

### HACIA EL RECONOCIMIENTO AUTOMÁTICO

Un corpus aleatorio de 2.160 preguntas fue clasificado por dieciséis alumnos de Ingeniería Civil Informática en la asignatura de Sistemas Inteligentes, familiarizados con este tipo de análisis y minería de datos, en base a una taxonomía de cinco categorías de motivaciones. Con esto, los investigadores exploraron una amplia variedad de estrategias de clasificación multietiqueta para construir modelos algorítmicos que reconocieran los rasgos lingüísticos que caracterizan las motivaciones subyacentes. Asimismo, hallaron que mientras más compleja es la sintaxis de una oración, más motivaciones se expresan en la pregunta.

De acuerdo a Alejandro Figueroa, investigador del Departamento de Ciencias de la Ingeniería de la UNAB, este trabajo está en la frontera de la lingüística computacional, psicología y análisis de datos. "Lo que abordamos es un problema complejo por definición, ya que los problemas tradicionales consisten en asignar una etiqueta a un objeto; sin embargo, en este caso, el número de etiquetas es indeterminado. Lo que descubrimos es que tomar en cuenta la correlación de las etiquetas ayuda a mejorar los resultados, principalmente, porque hay temas particulares donde suelen unirse algunas motivaciones", destaca el académico. Publicada este año en el journal internacional Expert Systems with Applications, esta investigación fue realizada gracias al financiamiento del proyecto FONDECYT 11130094, y desarrollada en el marco de una tesis del Magíster en Ciencias de la Computación (MCC) de la UNAB, en su línea de análisis de datos.