

Generación automática de palabras clave para monitorizar dominios en redes sociales

Miguel Ángel Rozalén Soriano, Universidad Carlos III de Madrid, Knowledge Reuse Group (España), ORCID: 0000-0002-5897-9018.

Álvaro Aranda Jiménez, Universidad Carlos III de Madrid, Knowledge Reuse Group (España), ORCID: 0000-0003-3859-1574.

Resumen

En la última década las redes sociales se han convertido en uno de los pilares básicos de los canales de comunicación en el cual sus usuarios participan generando contenido de todo tipo e interactuando entre ellos. Esta situación proporciona de forma continua una ingente cantidad de información que se puede analizar para evaluar su impacto y su posible aplicación a multitud de ámbitos.

En el mundo de los medios de comunicación, esta información proveniente de usuarios de redes sociales puede ser una fuente directa de generación de noticias, incluyendo, además de textos, imágenes o videos. Se puede estar alterando así el flujo de la noticia, provocando su nacimiento en las redes sociales. Es cuando, los medios se enfrentan a una oportunidad de aprovechar esa valiosa información generada en tiempo real.

Para poder identificar este tipo de sucesos en tiempo real desde las redes sociales, serán necesarias herramientas sobre las que apoyarse para poder monitorizar las mismas. Difícilmente va a ser posible escuchar, o más bien entender, todo lo que sucede a la vez, pero sí surge la posibilidad de monitorizar sucesos para dominios concretos previamente definidos.

Es en este punto donde nace el objetivo tratado en esta publicación: ser capaz de generar la información necesaria, y de manera automática, para poder “escuchar” las redes sociales como monitorización de un dominio concreto. Teniendo presente, además, que, aunque todos los recursos puedan salir bajo un mismo contexto, se puede hablar del mismo concepto de formas muy diversas, más en un lenguaje tan abierto e interpretable como es el castellano, y siendo conscientes que no se escribe de igual manera en medios de comunicación que en redes sociales, donde cada usuario puede escribir libremente, incluso con faltas de ortografía o generando nuevo vocabulario.

El caso está basado para el uso de la red social Twitter. En el proceso de ‘escucha’ para la obtención de los recursos que se generan en dicha red, se le debe indicar hasta un máximo de 400 palabras que deben ser seguidas. La obtención de esas palabras que se le indican a la red social son clave para obtener resultados de calidad.

En esta comunicación se describirá el procedimiento llevado a cabo para maximizar la cantidad de palabras de calidad con las que se pueda poner a ‘escuchar’ la red social con el objetivo de obtener la mayor cantidad de recursos útiles para un dominio de interés.

Teniendo como concepto básico la generación y expansión de palabras clave, y la posible aplicación a cualquier otro dominio.

El caso está validado en colaboración con RTVE, con quien se puso en marcha un sistema de escucha con el que se pudo evaluar el funcionamiento. Se escogió como dominio de monitorización ‘accidentes de tráfico’, ya que implican directamente a personas y, desafortunadamente, suceden con una frecuencia alta. Se evaluó si el sistema de monitorización de la red social había registrado el accidente o no, y en caso afirmativo con cuánta diferencia horaria lo había hecho respecto a la noticia.