

## Resumen

El Procesamiento del Lenguaje Natural (PLN) es un campo de la informática, la inteligencia artificial y la lingüística computacional centrado en las interacciones entre las máquinas y el lenguaje de los humanos. Uno de sus mayores desafíos implica capacitar a las máquinas para inferir el significado del lenguaje natural humano. Con este propósito, diversas representaciones del significado y el contexto han sido propuestas obteniendo un rendimiento competitivo. Sin embargo, estas representaciones todavía tienen un margen de mejora en escenarios transdominios y translingües.

En esta tesis estudiamos el uso de grafos de conocimiento como una representación transdominio y translingüe del texto y su significado. Un grafo de conocimiento es un grafo que expande y relaciona los conceptos originales pertenecientes a un conjunto de palabras. Sus propiedades se consiguen gracias al uso como base de conocimiento de una red semántica multilingüe de amplia cobertura. Esto permite tener una cobertura de cientos de lenguajes y millones de conceptos generales y específicos del ser humano.

Como punto de partida de nuestra investigación empleamos características basadas en grafos de conocimiento – junto con otras tradicionales y meta-aprendizaje – para la tarea de PLN de clasificación de la polaridad mono- y transdominio. El análisis y conclusiones de ese trabajo muestra evidencias de que los grafos de conocimiento capturan el significado de una forma independiente del dominio. La siguiente parte de nuestra investigación aprovecha la capacidad de la red semántica multilingüe y se centra en tareas de Recuperación de Información (RI). Primero proponemos un modelo de análisis de similitud completamente basado en grafos de conocimiento para detección de plagio translingüe. A continuación, mejoramos ese modelo para cubrir palabras fuera de vocabulario y tiempos verbales, y lo aplicamos a las tareas translingües de recuperación de documentos, clasificación, y detección de plagio. Por último, estudiamos el uso de grafos de conocimiento para las tareas de PLN de respuesta de preguntas en comunidades, identificación del lenguaje nativo, y identificación de la variedad del lenguaje.

Las contribuciones de esta tesis ponen de manifiesto el potencial de los grafos de conocimiento como representación transdominio y translingüe del texto y su significado en tareas de PLN y RI. Estas contribuciones han sido publicadas en diversas revistas y conferencias internacionales.