

# Geographic Information Retrieval

Fernando S. Peregrino, David Tomás and Fernando Llopis

Departamento de Lenguajes y Sistemas Informáticos, Universidad de Alicante  
Carretera San Vicente del Raspeig s/n - 03690 Alicante (España)

**Resumen** Nuestra investigación se centra en la rama de la recuperación de información geográfica, concretamente se está investigando sobre las entidades geográficas vagas tales como “el norte de España”, “cerca de la costa mediterránea”, etc. Hasta la fecha solamente hemos hecho una aproximación a este tema, la cual fue hecha desde un punto de vista geográfico, más que de PLN. Para dicho artículo se desarrolló una aplicación que construía una imagen ráster o bitmap de la entidad geográfica concisa que aparecía en una consulta. Sobre esta imagen se asignaban valores a cada una de las cuadrículas que la formaban dando un mayor valor a aquellas que se encontraban más próximas a la entidad vaga que en la consulta se solicitaba. A modo de ejemplo, en el artículo se mostraron consultas del estilo de “Estaciones de esquí en el norte de España”, de la cual, el sistema obtenía la imagen raster de España (la entidad concisa) y daba un mayor peso a las cuadrículas que estaban en el norte (la expresión vaga) de esta entidad concisa. En definitiva, centrándonos en el ejemplo anterior, la imagen ráster nos permitía ponderar la consulta realizada (Estaciones de esquí) mediante la entrada como parámetro de las coordenadas geográficas de cualquier estación de esquí del mundo, discriminando fácilmente el sistema aquellas que estaban en el norte de España del resto, incluso las que estaban muy próximas en la frontera de la entidad difusa como podían ser las estaciones de Andorra, Pirineos franceses o incluso las del sistema central español. Tras este enfoque pretendemos abordar la investigación de las entidades geográficas vagas mediante la captación de información en lenguaje natural en Internet, concretamente en páginas de agencias de viaje, turismo, opinión, redes sociales, tales como Booking, TripAdvisor, Flickr etc., y con ayuda de la Wikipedia.

Por otra parte, si lo expuesto anteriormente es el trabajo que hemos realizando y en el que nos vamos a centrar en un futuro próximo, hay que hablar de lo que actualmente estamos preparando, un sistema con el que vamos a participar en el Workshop del NTCIR, en la rama de GeoTime (recuperación de información geográfica y temporal), para lo cual se cuenta con un sistema de recuperación de información temporal (TIP-Sem) desarrollado en nuestro grupo, y se está implementando la parte geográfica mediante Lucene y el API de Yahoo! PlaceMaker. Por último, y como objetivo final, lo que se pretende es implantar un sistema GIR (Geographic Information Retrieval) acoplado la parte de recuperación geográfica general que vamos a presentar en el NTCIR y la específica de la entidades difusas, así como cualquier otro problema relacionado con la recuperación de información geográfica.

**Keywords:** Geographic Information Retrieval, Fuzzy Entities, Vernacular Language.