



Migración de un framework para la búsqueda por similitud en espacios métricos

Karina Mariela Figueroa Mora¹ y Yashar Linares²

¹ Fac. de cs físico-matemáticas, ² Universidad Michoacana de San Nicolás de Hidalgo. karina@fismat.umich.mx

La búsqueda por similitud consiste en recuperar de una base de datos los elementos mas parecidos a uno de consulta. Este tipo de búsqueda es la única alternativa en bases de datos multimedia pues en éstas es impensable pensar en una comparación por igualdad. Piense por ejemplo en una base de datos de imágenes satelitales. Una imagen tomada algún tiempo después no sería posible compararla por igualdad pues las imágenes en la base de datos seguro tendrán variaciones.

Dado que existe una gran variedad de elementos multimedia, en algunos la única manera de comparar sea con una distancia definida por expertos en el área. Este tipo de escenarios es factible manejarlos como un espacio métrico, pues éstos están conformados por una base de datos y una función de distancia.

En 2008 se hizo un esfuerzo por crear un framework en el que: fuera amigable y solo se trabajará en un algoritmo, la comunidad internacional lo usara para homogeneizar las medidas de comparación y; poder valorar y evaluar un nuevo algoritmo. En ese momento el framework de desarrollo se escribió en lenguaje c y fue una referencia en todo el mundo. A medida que ha avanzado el tiempo y que se han popularizado los lenguajes de programación dicho framework quedó obsoleto y para las nuevas generaciones la curva de aprendizaje es muy costosa. Sin embargo, para la comunidad internacional que trabaja en este tipo de búsquedas esta herramienta sigue siendo muy importante para poder comparar los algoritmos nuevos con los existentes.

La metodología usada fue la siguiente: primero se analizó qué del código anterior podía ser un objeto y se diseñó cada objeto resultante. Posteriormente se buscó ensamblar cada uno de los objetos por medio de objetos genéricos. La migración presentada implicó un amplio estudio del manejo de tipos en el lenguaje java. Una ventaja de este nuevo framework es que ahora será posible ejecutar algoritmos al mismo tiempo, sobre las mismas condiciones de la base de datos. Esto es factible aprovechando que java es un lenguaje orientado a objetos.

En este trabajo se presenta la metodología empleada para migrar el framework de desarrollo a un nuevo lenguaje con, incluso, un paradigma distinto, el lenguaje seleccionado fue java. De manera específica mostraremos las implementaciones hechas para tres de los algoritmos mas referenciados en esta comunidad. Como trabajo a futuro se espera poder liberar el código de este nuevo framework a la comunidad internacional y esperar robustecer este framework con las colaboraciones de los investigadores que trabajan en este tema.