

Título: OPTIMIZACIÓN DE BÚSQUEDAS EN GRANDES CONJUNTOS DE DATOS MEDIANTE LA UTILIZACIÓN DE ALGORITMOS DE CLUSTERING CON PRESERVACIÓN DE LA TOPOLOGÍA

Nombre: Quintana Pacheco, Yuri

Universidad: Universidad de Alicante

Departamento: Tecnología informática y computación

Fecha de lectura: 25/02/2013

Programa de doctorado: Tecnologías de la sociedad de la información

Dirección:

> **Director:** DANIEL RUÍZ FERNÁNDEZ

Tribunal:

> **presidente:** DIEGO ANDINA DE LA FUENTE

> **secretario:** PATRICIA COMPAÑ ROSIQUE

> **vocal:** FRANCISCO JAVIER GARCÍA CASADO

Descriptor:

> HEURISTICA

El fichero de tesis ya ha sido incorporado al sistema

> <http://hdl.handle.net/10045/28507>

Resumen: En la presente investigación se propone un método para realizar búsquedas en grandes conjuntos de datos con características similares a los grafos. El método está dividido en dos etapas: una etapa de organización del conjunto de datos y una etapa donde se puede realizar la búsqueda de elementos que minimicen una función objetivo según el modelo organizativo obtenido.

Para la organización del conjunto de datos se propone el empleo del algoritmo Growing Neural Gas, al que se le han incorporado modificaciones que facilitan su empleo en el contexto de la investigación. El método de grafos embebidos en espacios vectoriales se utiliza para obtener representaciones de los elementos del conjunto de datos apropiadas para la aplicación del algoritmo de clustering. Como resultado del proceso de organización cada elemento del conjunto de datos es asociado a un nodo del modelo obtenido de la aplicación del algoritmo de clustering.

Para la etapa de búsqueda se propone un algoritmo basado en la heurística Simulated Annealing. El algoritmo realiza estimaciones de la calidad de los elementos asociados a cada nodo del modelo mediante evaluaciones de la función objetivo que realiza mientras se recorre el modelo. Estas estimaciones son empleadas para orientar la búsqueda.

El funcionamiento de cada etapa de la propuesta de solución es validada de forma empírica mediante el empleo

de conjuntos de datos y funciones objetivo seleccionados con este fin. Los resultados de la aplicación del método general propuesto muestran que la obtención de soluciones requiere la evaluación de un pequeño porcentaje de los elementos del conjunto de datos, de modo que se reducen los tiempos asociados al proceso de búsqueda.