

**RESUMEN DE PROPUESTA DE TOG**

**INTEGRACIÓN DE UN ALGORITMO DE SUMARIZACIÓN DE GRAFOS CON UNA BASES DE DATOS  
DE GRAFOS (NEO4J)**

Proponente: Mtro. Víctor Ortega, ITESO/UABCC, [vortega@iteso.mx](mailto:vortega@iteso.mx)

**1) Descripción general del proyecto propuesto**

Los grafos son de gran utilidad para el análisis de información semiestructurada. Una de las aplicaciones es para el análisis de redes sociales, sin embargo, cuando la información es demasiada, resulta casi imposible realizar un análisis de lo que está ocurriendo al interior. La sumarización de un grafo ayuda en el proceso de descubrimiento.

En el primer semestre de 2018, el doctor Luis Fernando Gutiérrez, el doctor Francisco Cervantes y un servidor implementados un algoritmo para sumarizar grafos que toma en cuenta los atributos para generar agrupaciones. El algoritmo toma como entrada archivos CSV, con la información de los nodos, atributos y relaciones y está codificado en java.

El objetivo del proyecto es realizar la conexión directa con la base de datos de grafos (NEO4J), para que mediante de ejecución de *querys* escritos en el lenguaje de consultas nativo de esta base de datos (CYPHER) el algoritmo obtenga directamente las entradas de la base de datos basada en grafos y no exista la necesidad de utilizar archivos CSV.

Las principales etapas del proyecto son:

- 1.1 Entender el algoritmo base
- 1.2 Entender cómo está implementado en java
- 1.3 Desarrollar el código necesario en java para que la implementación del algoritmo reciba de información directamente de NEO4J a través de querys de CYPHER
- 1.4 Escribir un artículo con los resultados y el trabajo de obtención de grado (TOG)

**2) Vinculación o Colaboración**

Se trabajará de forma indirecta con el Dr. Francisco Cervantes y el Dr. Luis Fernando Gutiérrez

**3) Asignaturas de la MSC relacionadas con el desarrollo del proyecto**

- 3.1. Matemáticas Avanzadas para Computación
- 3.2. Análisis y Diseño de Algoritmos
- 3.3. Bases de Datos Avanzadas
- 3.4. Manejo y Análisis de Información Masiva



**ITESO**  
Universidad Jesuita  
de Guadalajara

**RESUMEN DE PROPUESTA DE TOG**

**4) Estudiante participante en la propuesta**

Dos estudiantes

**5) Bibliografía relacionada**

- 5.1 D. Koutra, U. Kang, J. Vreeken, and C. Faloutsos, "Summarizing and understanding large graphs," *Stat. Anal. Data Min.*, vol. 8, no. 3, pp. 183-202, Jun 2015. [Online]. Available: <http://dx.doi.org/10.1002/sam.11267>
- 5.2 K. U. Khan, W. Nawaz, and Y.K. Lee, "Set-Based Unified Approach for Attributed Graph Summarization," in (BdCloud) 2014 IEEE Fourth Int. Conf. on. Big Data and Cloud Computing, Sydney, NSW, Dec. 2014, pp.378-385.
- 5.3 T. Yuanyuan, R. A. Hankins, and J. M. Patel, "Efficient aggregation for graph summarization," in *Proceedings of the 2008 ACM SIGMOD Int. Conf. on Management of Data*, Vancouver, Canada, Jun. 2008, pp. 567-580.
- 5.4 NEO4J. (2015). Sustainable Competitive Advantage: Creating Business Value through Data Relationships [Online]. Available: [https://info.neo4j.com/rs/neotechnology/images/wp\\_sca\\_neo4j.pdf](https://info.neo4j.com/rs/neotechnology/images/wp_sca_neo4j.pdf).
- 5.5 O. Panzarino, *Learning Cypher*. Birmingham, UK: Packt Publishing Ltd, 2014.

**6) Víctor H. Ortega Guzmán**

Maestro en Computación (UNIVA) con especialidad en Tecnologías para *Business Intelligence: Data warehouse y Data Mining* (UOC, España) y otra en Mejora de Procesos de Negocio (ITESO). Consultor en el Centro de Gestión de la Innovación y la Tecnología (Ceginnt). Trabaja en proyectos relacionados con las áreas de bases de datos, *business intelligence*, y bases de datos basadas en grafos. Actualmente es estudiante del Doctorado en Ciencias de la Ingeniería en la línea de Software de Alto Desempeño (ITESO).