
Formulario de Aprobación Curso de Actualización

Asignatura: Seminario de Bases de Datos de Grafos

(Si el nombre contiene siglas deberán ser aclaradas)

Profesor de la asignatura ¹: Dr. Alejandro A. Vaisman, Instituto Tecnológico de Buenos Aires
(título, nombre, grado o cargo, Instituto o Institución)

Profesor Responsable Local ¹: Dra. Adriana Marotta, grado 4, Instituto de Computación
(título, nombre, grado, Instituto)

Otros docentes de la Facultad:
(título, nombre, grado, Instituto)

Docentes fuera de Facultad:
(título, nombre, cargo, Institución, país)

Instituto ó Unidad: Instituto de Computación
Departamento ó Area:

¹ Agregar CV si el curso se dicta por primera vez.
(Si el profesor de la asignatura no es docente de la Facultad se deberá designar un responsable local)

Horas Presenciales: 20

Público objetivo y Cupos: Profesionales Informáticos y áreas afines interesados en el área de Manejo de Datos e Información.

Objetivos: (1) Presentar los fundamentos de los sistemas de gestión de bases de datos de grafos (GDBMS). Entender los modelos de datos en los que se basan, su implementación, y su uso en el contexto actual de "Big Data". Distinguir los problemas para los cuales utilizar grafos es más eficiente que utilizar los sistemas tradicionales. (2) Presentar y discutir los temas de investigación y problemas abiertos. (3) Interesar a los participantes a integrar proyectos de investigación en el tema. Se discutirán los dos tipos de grafos utilizados actualmente: property graphs y RDF triple stores. Se caracterizarán y estudiarán diferentes tipos de consultas sobre BD de grafos. Luego se profundizará sobre property graphs, y su implementación más difundida, Neo4j.

Conocimientos previos exigidos: Conocimientos de Bases de Datos y SQL

Conocimientos previos recomendados:

Metodología de enseñanza:

(comprende una descripción de la metodología de enseñanza y de las horas dedicadas por el estudiante a la asignatura, distribuidas en horas presenciales -de clase práctica, teórico, laboratorio, consulta, etc.- y no presenciales de trabajo personal del estudiante)

El curso consistirá en dos partes: durante las primeras clases se brindarán los fundamentos, y luego se trabajará con tres modelos como ejemplo: las bases de datos Northwind, Movies, y Worldcup.

Para aprobar el curso, se requerirá desarrollar un set de consultas sobre cada una de las tres bases de datos de ejemplo, y desarrollar un modelo sobre un problema a definir. El primer proyecto será individual, el segundo grupal, en grupos de a lo sumo 3 estudiantes.

- Horas clase (teórico-práctico): 12
- Horas clase (práctico):
- Horas clase (laboratorio):
- Horas consulta: 8
- Horas evaluación:
 - Subtotal horas presenciales: 20
- Horas estudio:
- Horas resolución ejercicios/prácticos:
- Horas proyecto final/monografía: 25
 - Total de horas de dedicación del estudiante: 45

Forma de evaluación:

- Realización de dos proyectos, uno individual y otro grupal.

Temario:

1. Contexto. Conceptos de Big Data. El paradigma NoSQL.
2. Modelos de datos. Historia. El modelo de grafos:
 - Modelo de grafos como alternativa al modelo relacional.
 - Casos de uso: biología, web, redes sociales.
 - Property graphs vs RDF triple stores.
 - BD de grafos vs. frameworks para procesamiento de grafos.
3. El TDA grafo. Operaciones. Implementaciones
4. Tipos de consultas sobre grafos: patrones y consultas navegacionales.
5. Property graphs. Modelo formal.
6. Property graphs. Implementaciones.
7. El modelo de grafos edge-labeled: RDF
8. Consultando RDF: SPARQL.
9. Implementaciones del modelo RDF - triple stores.
10. Frameworks de procesamiento de grafos.
11. Modelado y consultas sobre property graphs utilizando Neo4j.
12. Cypher: el lenguaje de consulta para Neo4j.
13. Consultando BDG con Cypher.

Bibliografía:

(título del libro-nombre del autor-editorial-ISBN-fecha de edición)

- [1] R. Angles. A Comparison of Current Graph Database Models. In Proceedings of ICDE Workshops, pages 171–77, Arlington, VA, USA, 2012.
- [2] Renzo Angles, Marcelo Arenas, Pablo Barceló, Aidan Hogan, Juan L. Reutter, and Domagoj Vrgoc. Foundations of modern graph query languages. CoRR, abs/1610.06264, 2016.

-
- [3] Renzo Angles and Claudio Gutierrez. Survey of graph database models. *ACM Comput. Surv.*, 40(1):1:1–1:39, 2008.
- [4] NoSQL Databases. <http://nosql-database.org/>.
- [5] Apache Giraph. <http://giraph.apache.org/>.
- [6] O. Hartig. Reconciliation of RDF* and property graphs. *CoRR*, abs/1409.3288, 2014.
- [7] Huahai He and AmbujK. Singh. Query language and access methods for graph databases. In *Managing and Mining Graph Data*, volume 40 of *Advances in Database Systems*, pages 125–160. Springer US, 2010.
- [8] Grzegorz Malewicz, Matthew H. Austern, Aart J.C Bik, James C. Dehnert, Ilan Horn, Naty Leiser, and Grzegorz Czajkowski. Pregel: a system for large-scale graph processing. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, pages 135–146. ACM, 2010.
- [9] Neo4j. <http://neo4j.org/>.
- [10] N. U. Rehman, A. Weiler, and M. H. Scholl. OLAPing social media: the case of twitter. In *Advances in Social Networks Analysis and Mining 2013, ASONAM '13*, pages 1139–1146, Niagara, ON, Canada, 2013.
- [11] Ian Robinson, Jim Webber, and Emil Eifrem. *Graph Databases*. O'Reilly Media, Inc., 2013.
- [12] Semih Salihoglu and Jennifer Widom. GPS: A Graph Processing System. Technical report, Stanford University, 2012.
- [13] Sabri Skhiri and Salim Jouili. Large graph mining: Recent developments, challenges and potential solutions. In Marie-Aude Aufaure and Esteban Zim'anyi, editors, *Business Intelligence*, volume 138 of *Lecture Notes in Business Information Processing*, pages 103–124. Springer Berlin Heidelberg, 2013.
- [14] Srinath Srinivasa. *Data, Storage and Index Models for Graph Databases*, page 47. Information Science Reference, 2011.
- [15] Yuanyuan Tian, Richard A. Hankins, and Jignesh M. Patel. Efficient aggregation for graph summarization. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*, pages 567–580. ACM, 2008.
- [16] Titan. <http://thinkaurelius.github.com/titan/>.
- [17] Chad Vicknair, Michael Macias, Zhendong Zhao, Xiaofei Nan, Yixin Chen, and Dawn Wilkins. A comparison of a graph database and a relational database: a data provenance perspective. In *Proceedings of the 48th Annual Southeast Regional Conference*, pages 42:1–42:6. ACM, 2010.
-

Datos del curso

Fecha de inicio y finalización: Desde el 25/3/2019 hasta el 12/4/2019

Horario y Salón: Lunes a jueves en semana del 25 de marzo al 28 de marzo. Siguiendo dos semanas a distancia.
Horario y Salón: a definir

Arancel: \$9.810
