

**MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN
PROPUESTA DE CURSO DE POSGRADO**

1- DATOS GENERALES DE LA ACTIVIDAD CURRICULAR	
1.1 Título del Curso	Generación de Grafos de Conocimiento para Explorar Silos de Datos: Fundamentos, Tecnologías y Aplicaciones
1.2 Área temática ¹	Ingeniería de Conocimiento - Sistemas de Inteligencia Artificial - Sistemas de Información - Web Semántica

2- COMPOSICIÓN DEL EQUIPO DOCENTE	
2.1 Responsable a cargo de la actividad curricular	Germán A. Braun - Laura A. Cecchi
2.2 Docentes	

3- CARGA HORARIA					
Carga horaria teórica	30 horas				
Carga horaria práctica	30 horas				
Carga horaria total	60 horas				
Distribución horaria semanal	Lu	Ma	Mie	Jue	Vie
Fecha de inicio sugerida					

¹ Corresponde a uno de los siguientes tópicos: Algoritmos y Lenguajes; Teoría de la Computación; Ingeniería de Software, Bases de Datos y Sistemas de Información; Arquitecturas, Sistemas Operativos y Redes.

4- BREVE RESUMEN DE CONTENIDOS (hasta 400 palabras)

En años recientes, los grafos de conocimiento han alcanzado una atención significativa, en la industria y en la academia, como un medio para estructurar datos y conocimiento sobre el mundo real, a partir de fuentes de datos diversas. El concepto fue desarrollado en sus inicios por Google, a través del “Google Knowledge Graph”, con el objetivo de mejorar las capacidades semánticas de sus búsquedas: usando “entidades” en lugar de “cadenas de caracteres”.

Sin embargo, el concepto fue adoptado ampliamente con el surgimiento de la Web Semántica y, en particular, la estandarización de lenguajes para representación de grafos de datos como RDF, los lenguajes de consulta como SPARQL y las tecnologías para la gestión de bases de datos de grafos.

En la práctica, los grafos de conocimiento han sido probados como una abstracción útil y escalable para representar, integrar y explorar fuentes de datos heterogéneas, en una representación unificada. Asimismo, se han desarrollado exitosas aplicación basadas en grafos, muchas de las cuales pueden ser accedidas libremente como Wikidata, Yago, entre otras.

En este curso, se introducen los conceptos fundamentales referidos a los grafos de conocimiento, presentando los fundamentos teóricos y lenguajes de modelado y consulta. Asimismo, se presentan ejemplos prácticos del uso, métodos y herramientas para la generación de dichos grafos.

5- CONOCIMIENTOS PREVIOS REQUERIDOS

Nociones de modelado de datos, lenguajes (UML, ER)

Nociones y práctica en el uso de bases de datos relacionales y no relacionales

Conocimientos de lógica de primer orden a nivel de cursos introductorios de pregrado.

Nociones básicas en Python y/o Java

6- OBJETIVOS

Los objetivos del curso son:

1. Introducir al estudiante en la gestión de grafos de conocimiento, enfocándonos en sus fundamentos y en las tecnologías existentes para generarlos, mantenerlos y utilizarlos.
2. Presentar lenguajes para la representación de datos basado en grafos (RDF) y de ontologías (OWL, SHACL) y desarrollar habilidades para su buen uso.
3. Estudiar métodos y herramientas para la generación y validación de grafos de conocimiento.
4. Plantear soluciones basadas en grafos de conocimiento determinando y aplicando las herramientas adecuadas
5. Explorar posibles perspectivas para el futuro, teniendo en cuenta los conceptos, ideas y herramientas de los grafos de conocimiento.

En este sentido, se estudiarán los lenguajes de representación basados en grafos estandarizados por W3C, y su relación con formalismos clásicos, de modo que el estudiante pueda desarrollar habilidades básicas para la creación de sistemas basados en conocimiento. Asimismo, estas nociones se relacionarán con otras áreas de las Ciencias de la Computación, tales como lenguajes formales, bases de datos, Ingeniería de Software, Inteligencia Artificial, etc. en los cuales los grafos de conocimiento encuentran sus fundamentos.

7- CONTENIDOS (organizados en unidades, ejes, módulos, otros)

Unidad I: Introducción

Introducción a los Grafos de Conocimiento: definición, antecedentes y usos. Introducción a la Web Semántica y a las ontologías. Grafos de Conocimiento en la Web de Datos.

Unidad II: Lenguajes

Grafos en RDF. Ontologías RDFS y OWL. Namespaces. URIs e IRIs. Lenguajes de consulta: SPARQL. Federación. Validación: SHACL.

Unidad III: Creación de Grafos de Conocimientos

Modelos de Grafos de Conocimiento. Metodologías. Integración de datos. Virtualización y Materialización. Creación de grafos a partir de datos heterogéneos: bases de datos relacionales, mapeos R2RML/RML, hojas de cálculos, datos estructurados.

Unidad IV: Tecnologías y Herramientas para Grafos de Conocimiento

Vocabularios: foaf, schema.org, dcat. Motores para almacenar grafos RDF: GraphDB, Stardog. Motores para la generación de grafos de conocimiento a partir de datos heterogéneos: morph-kgc, ontop, OntoRefine. Herramientas para manipular y consultar datos semánticos: RDFlib. Consultas a Grafos de Conocimiento aplicando Inteligencia Artificial generativa.

Unidad V: Aplicación basadas en Grafos

Grafos de Conocimiento en el mundo real: Google, DBpedia, YAGO, WordNet, Wikidata.

8- PROPUESTA DIDÁCTICA (metodología de trabajo de clases teóricas y prácticas)

La propuesta metodológica consiste en la combinación de clases teóricas donde se introducen los conceptos fundamentales del curso y de ejercitación mediante trabajos prácticos grupales de los conceptos presentados a nivel teórico.

Las clases teóricas son conducidas por el profesor quien desarrollará los temas mediante una exposición oral, con la ayuda de algún recurso didáctico visual. Al inicio de la clase teórica se hará una síntesis o revisión de los conceptos y resultados a tener presentes para el desarrollo de la misma. Los temas serán introducidos a partir de ejemplos motivadores que muestren su necesidad. Además, se presentarán, siempre que sea posible, ejercicios simples para reforzar el concepto explicado y vincularlo con otras áreas de las Ciencias de la Computación.

Para un buen aprendizaje es necesario que el alumno pueda participar activamente, ya sea en forma individual o grupal, y discutir los temas desarrollados. Por tal motivo es que las clases teóricas no consistirán en una excesiva exposición de temas y resultados, sino que con el fin de propiciar su comprensión, se estimulará el pensamiento reflexivo y se promoverá la participación de los alumnos a través de preguntas y trabajos sencillos.

En las clases prácticas el rol de los alumnos es más activo y los docentes brindan pautas que orientan la resolución de los ejercicios y problemas propuestos en los trabajos prácticos. Los ejercicios y problemas serán seleccionados con el objetivo de que, al resolverlos, los alumnos puedan familiarizarse con el uso de los lenguajes de grafos, con sus fundamentos teóricos y ganar experiencia con las tecnologías relacionadas. Se utilizarán bases de datos de grafos y herramientas para visualizar e interactuar con grafos de conocimiento.

9- MODALIDAD DE EVALUACIÓN Y CONDICIONES DE ACREDITACIÓN²

Evaluación: para el seguimiento de los estudiantes se registrará la asistencia y la participación en clases. La metodología de evaluación consiste del desarrollo y entrega de actividades prácticas propuestas en clase. Asimismo, incluye la elaboración de un trabajo final de investigación individual o en grupos u opcionalmente, podrá instrumentarse un examen final escrito abordando contenidos teóricos y prácticos.

Acreditación: 80% de asistencia. Realización, entrega y aprobación de las actividades prácticas con nota superior o igual a 7 (siete) puntos. Elaboración del trabajo final o bien desarrollo del examen final; en ambos casos, la nota debe ser no menos de 7 (siete) puntos.

² Son condiciones mínimas para la aprobación de todos los cursos: cumplir con un mínimo del 80% de asistencia a las clases, realizar las tareas y aprobar las evaluaciones que se hayan propuesto en el programa, con una calificación no menor a 7 (puntos). Los trabajos de evaluación pautados y la calificación de los alumnos deberán realizarse dentro de los 60 días posteriores a la finalización del curso.

10-

BIBLIOGRAFÍA DE LECTURA OBLIGATORIA CORRESPONDIENTE A CADA UNIDAD Y GENERAL

Bibliografía:

- **The Semantic Web.** Berners-Lee, Tim, James Hendler, and Ora Lassila. Scientific American. (2001)
- **Knowledge graphs.** Sequeda, Gutierrez. Commun. ACM (2021)
- **Knowledge Graphs.** Aidan Hogan, Eva Blomqvist, Michael Cochez, Claudia d'Amato, Gerard de Melo, Claudio Gutierrez, Sabrina Kirrane, José Emilio Labra Gayo, Roberto Navigli, Sebastian Neumaier, Axel-Cyrille Ngonga Ngomo, Axel Polleres, Sabbir M. Rashid, Anisa Rula, Lukas Schmelzeisen, Juan F. Sequeda, Steffen Staab and Antoine Zimmermann. *ACM Computing Surveys (CSUR)* 54, no. 4:1-37. (2022)
- **Virtual Knowledge Graphs: An Overview of Systems and Use Cases.** Guohui Xiao, Linfang Ding, Benjamin Cogrel, Diego Calvanese. Data Intell. (2019)
- **Morph-KGC: Scalable knowledge graph materialization with mapping partitions.** Arenas-Guerrero et. al. Semantic Web. (2022)
- **A Developer 's Guide to the Semantic Web,** Liyang Yu. Springer; 2nd ed. Corr. 3rd printing edition, 2015 ISBN: 978-3662437957 (2014)
- **Semantic Web for the Working Ontologist: Effective Modeling for Linked Data, RDFS, and OWL.** Dean Allemang, James Handler and Fabien Gardon. 3rd edition. ACM Books. (2019)
- **Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges.** Illaria Tiddi, Freddy Lécué and Pascal Hitzler. IOS Press. (2020)
- **Ontop: Answering SPARQL Queries over Relational Databases.** Diego Calvanese, Benjamin Cogrel, Sarah Komla-Ebri, Roman Kontchakov, Davide Lanti, Martin Rezk, Mariano Rodriguez-Muro, and Guohui Xiao. Semantic Web Journal. (2017)
- **Efficient SPARQL-to-SQL with R2RML mappings.** Mariano Rodriguez-Muro and Martin Rezk, J. of Web Semantics. (2015)
- **Linked Data - The Story So Far.** Bizer, C., Heath, T. and Berners-Lee, T. In *Semantic services, interoperability and web applications: emerging concepts* (pp. 205-227). IGI global. (2011)
- **XML Schema Definition Language (XSD) 1.1. World Wide Web Consortium** <https://www.w3.org/TR/xmlschema11-1/>
- **OWL 2. W3C Recommendation.** <http://www.w3.org/TR/owl2-overview/>
- **R2RML: RDB to RDF Mapping Language. World Wide Web Consortium Recommendation.** <https://www.w3.org/TR/r2rml/>
- **Shapes Constraint Language (SHACL). World Wide Web Consortium Recommendation.** <https://www.w3.org/TR/shacl/>
- **SPARQL 1.1 Query Language. World Wide Web Consortium Recommendation.** <http://www.w3.org/TR/sparql11-query/>
- **Extensible Markup Language (XML). World Wide Web Consortium** <https://www.w3.org/TR/xml/>
- **RDF 1.1 Primer. World Wide Web Consortium.** <https://www.w3.org/TR/rdf11-primer/>

Direcciones útiles:

- **URI clarification.** <http://www.w3.org/TR/uri-clari#cation/>
- **URI schemes** <http://www.iana.org/assignments/uri-schemes> (IA-NA)
- **Ontop.** <https://ontop-vkg.org/>
- **schema.org.** <https://schema.org/>
- **foaf.** <http://xmlns.com/foaf/0.1/>
- **dcat.** <https://www.w3.org/TR/vocab-dcat-2/>

- **Stardog.** <https://www.stardog.com/>
- **Morph-KGC.** <https://morph-kgc.readthedocs.io/en/latest/>
- **GraphDB.** <https://graphdb.ontotext.com/>
- **OntoRefine.** <https://www.ontotext.com/products/ontotext-refine/>
- **DBpedia.** <https://www.dbpedia.org/>
- **YAGO.** <https://yago-knowledge.org/>
- **Wikidata.** <https://www.wikidata.org/>
- **RDFlib.** <https://rdflib.readthedocs.io/en/stable/>

11- INFRAESTRUCTURA E INSUMOS REQUERIDOS³

Laboratorio de Informática solamente para las clases prácticas. Cañón y pantalla solamente para las clases teóricas.

12 – OTRA INFORMACIÓN RELEVANTE

³ Deberá constar aquí si la realización del curso requiere contar con instalaciones especiales (laboratorio, sala de informática, equipamiento audiovisual, etc). Explicitar si se estima que el curso debe tener un número máximo determinado de asistentes para poder ser dictado.