

Formulario de Aprobación Curso de Posgrado 2010

Asignatura: Web de Datos

Profesor de la asignatura: Dr. Claudio Gutierrez, Dpto. de Ciencias de la Computación, Universidad de Chile

(título, nombre, grado o cargo, Instituto o Institución)

Profesor Responsable Local: Regina Motz, Grado 5, Instituto de Computación

Instituto ó Unidad: Instituto de Computación

Departamento ó Area: Grupo de Concepción de Sistemas de Información (CSI)

Fecha de inicio y finalización: 22/11/2010 al 26/11/2010

Horario y Salón: 17:00 a 20 horas (Salón a confirmar.)

Horas Presenciales: 15 hrs presenciales + 15 hrs tareas = 30 hrs.

Nº de Créditos: 2

Público objetivo y Cupos:

Estudiantes de Posgrado y Profesionales interesados en la publicación de datos abiertos y en el desarrollo de sistemas de información para la web semántica.

Objetivos:

El objeto de información tradicional ha sido el documento. Gradualmente, debido a diferentes avances tecnológicos, un objeto más básico se ha puesto al centro: los datos. Y de una Web de documentos (una "biblioteca universal") se ha pasado hoy a una Web de Datos, comenzando a realizar una "base de datos global".

Los objetivos de este breve curso son 1) Presentar los fundamentos de esta revolución en la creación, manejo y procesamiento y consumo de datos que se ha comenzado a producir. 2) Discutir el rol que estos datos juegan en el sistema de información actual, y los desafíos técnicos, sociales y legales que ello conlleva. 3) Introducir algunos de los modelos, técnicas y estándares que se están usando para enfrentar estos desafíos.

Conocimientos previos recomendados: Recomendado conocer alguna herramienta para el modelado de datos, como diseño relacional de bases de datos, MER o UML.

Metodología de enseñanza:

Clases teórico-prácticas en donde se incluirán discusiones sobre ejemplos de aplicación, modelado y ejercicios prácticos.

15 hrs presenciales de clase + 15 hrs de tarea domiciliaria = total 30 hrs.

Forma de evaluación:

Se prevee la realización de una tarea domiciliaria para la evaluación.

Temario:

1. Introducción / Publicación de datos.

- El rol que los datos juegan en el sistema de información actual. Desafíos (técnicos CS, sociales, etc.)
- Técnicas básicas usadas para desarrollar en la Web de datos: Identificadores, enlaces, metadatos.
- El lenguaje RDF. Ontologías
- Publicación de Datos Crudos

Ejercicio. Desarrollar modelo para codificar datos basado en triples (S,P,O), en formato NTriples y RDFa. Modelamiento usando conceptos como clases, propiedades e instancias.

2. Modelos de datos

- Web de documentos versus Web de datos
- Modelos de datos propuestos para conceptualizar la Web de datos y sus facetas.
- El proyecto Linked Data
- El movimiento Open Data

Ejercicio. A partir de datos reales, convertirlos en archivos HTML+RDFa para publicar en la Web, utilizando las técnicas presentadas.

3. Acceso a los datos

- El futuro del acceso a datos.
- Lenguajes de Consulta: SPARQL, endpoints.
- Grandes proyectos: data.gov

Ejercicio. Descargar la información publicada y consultarla usando SPARQL. Si tiempo (y background de estudiantes) permite, instalar y usar endpoints.

4. Procesamiento de datos

- Concepto, rol, alcances de técnicas de data mining y machine learning.
- Transformación de datos
- El "data scientist"

Ejercicio. Resolver problemas de procesamiento de datos definiendo consultas apropiadas para ello. Determinar técnicas necesarias.

5. Modelos de Infraestructura

- Límites del modelo relacional
- Arquitecturas alternativas
- Cloud

Facultad de Ingeniería

Comisión Académica de Posgrado

Universidad de la República – Facultad de Ingeniería, Comisión Académica de Posgrado/FING

Julio Herrera y Reissig 565, 11300 Montevideo, Uruguay

Tel: (+5982) 711-0544; Fax: (+5982) 711-5446 URL: <http://www.fing.edu.uy>

Ejercicio. Publicar y acceder a información utilizando alguna base de datos no-sql. Conocer la existencia de herramientas en la nube para procesar grandes volúmenes de datos.

6. Buenas prácticas de publicación

- Breve guía de buenas prácticas
- Análisis de un caso de estudio
- Evaluación

Ejercicio. Aplicar patrones de buen diseño a la publicación de datos. Analizar lo que ha publicado anteriormente en relación a los patrones de diseño usados.

Bibliografía:

The Claremont Report on Database Research, 2008.

Ch. Bizer et al., Linked Data - The Story So Far, 2009.

A. Halevy et al., Principles of dataspace systems, 2006

A. Mendelzon, T. Milo, Formal Models of Web Queries, 1998.

M. Stonebraker et al. The end of an architectural era: (it's time for a complete rewrite), 2007.

Tim O'Reilly, What Is Web 2.0, 2005.



Facultad de Ingeniería
Comisión Académica de Posgrado

2
nos

RDF Primer, <http://www.w3.org/TR/rdf-primer/>
No SQL, <http://nosql-database.org/>
DATA.gov project, <http://www.data.gov/>
