

# REPRESENTACIÓN DE LA INFORMACIÓN EN LA WEB. XML Y RECUPERACIÓN DE INFORMACIÓN

Curso 2020-2021

(Fecha última actualización: 30/04//2020)

(Fecha de aprobación en Comisión Académica del Máster: 01/07/2020)

SEMESTRE	CRÉDITOS	CARÁCTER	TIPO DE ENSEÑANZA	IDIOMA DE IMPARTICIÓN
1º	3	Optativa	Semipresencial	Español
<b>MÓDULO</b>		Recuperación de la Información		
<b>MATERIA</b>		Representación de la Información en la Web. XML y Recuperación de Información		
<b>CENTRO RESPONSABLE DEL TÍTULO</b>		Escuela Internacional de Posgrado		
<b>MÁSTER EN EL QUE SE IMPARTE</b>		<b>Máster Universitario en Información y Comunicación Científica</b>		
<b>CENTRO EN EL QUE SE IMPARTE LA DOCENCIA</b>		Facultad de Comunicación y Documentación		
<b>PROFESORES<sup>(1)</sup></b>				
<b>Eduardo Peis Redondo</b>				
<b>DIRECCIÓN</b>		Dpto. de Información y Comunicación, Facultad de Comunicación y Documentación. Despacho nº M Correo electrónico: epeis@ugr.es		
<b>TUTORÍAS</b>		<a href="https://directorio.ugr.es/static/PersonalUGR/*/show/1c333237007724c81e5fe07ebfef21c8">https://directorio.ugr.es/static/PersonalUGR/*/show/1c333237007724c81e5fe07ebfef21c8</a>		
<b>COMPETENCIAS GENERALES Y ESPECÍFICAS</b>				
<b>COMPETENCIAS BÁSICAS Y GENERALES</b>				
<p>CB6 - Poseer y comprender conocimientos que aporten una base u oportunidad de ser originales en el desarrollo y/o aplicación de ideas, a menudo en un contexto de investigación.</p> <p>CB7 - Que los estudiantes sepan aplicar los conocimientos adquiridos y su capacidad de resolución de problemas en entornos nuevos o poco conocidos dentro de contextos más amplios (o multidisciplinares) relacionados con su área de estudio</p> <p>CB8 - Que los estudiantes sean capaces de integrar conocimientos y enfrentarse a la complejidad de formular juicios a partir de una información que, siendo incompleta o limitada, incluya reflexiones sobre las responsabilidades sociales y éticas vinculadas a la aplicación de sus conocimientos y juicios</p> <p>CB9 - Que los estudiantes sepan comunicar sus conclusiones y los conocimientos y razones últimas que las sustentan a públicos especializados y no especializados de un modo claro y sin ambigüedades</p> <p>CB10 - Que los estudiantes posean las habilidades de aprendizaje que les permitan continuar</p>				

estudiando de un modo que habrá de ser en gran medida autodirigido o autónomo

### COMPETENCIAS ESPECÍFICAS

- CE20. - Conocer los nuevos métodos y técnicas de la recuperación de información.
- CE22. - Conocer la terminología, conceptos y aspectos fundamentales de las tecnologías relacionadas con la Web semántica.
- CE23. - Conocer la especificación XML y sus posibilidades de aplicación
- CE25. - Comprender el problema de la recuperación de información en el entorno de la Web
- CE26. - Analizar las posibilidades de mejora de la representación de recursos en la Web
- CE35. - Conocer y aplicar los fundamentos de los sistemas de recuperación de información y de los sistemas de recomendaciones

### OBJETIVOS O RESULTADOS DE APRENDIZAJE (SEGÚN LA MEMORIA DE VERIFICACIÓN DEL TÍTULO)

El alumno sabrá/comprenderá:

- Las posibilidades de mejora de la representación de recursos en la Web. Las limitaciones de HTML.
- La especificación XML (eXtensible Markup Language) y sus posibilidades de aplicación.
- La "familia" de especificaciones XML. Centrándose en XSL (eXtensible Stylesheet Language) para el estilo y procesamiento- y en RDF (Resource Description Framework) - para la lógica semántica-.
- El vocabulario para la recuperación de información estructurada NEXI
- Las posibles aplicaciones de las tecnologías en su conjunto: la Web Semántica y la Web de Datos.

El alumno será capaz de:

- Desarrollar una página Web XML con la ayuda de CSS y HTML sin utilizar editores
- Aplicar XSLT para modificar la "salida" de documentos estructurados XML
- Crear una proposición lógica utilizando RDF
- Recuperar información estructurada utilizando NEXI
- Mejorar las posibilidades de recuperación de información Web utilizando estas tecnologías

### BREVE DESCRIPCIÓN DE CONTENIDOS (SEGÚN LA MEMORIA DE VERIFICACIÓN DEL TÍTULO)

- Problemas de la recuperación de información en la Web y posibilidades de mejora en cuanto a la representación de recursos.
- Familia de especificaciones XML.
- Aplicaciones, desarrollos, viabilidad e impacto futuro.

### TEMARIO DETALLADO DE LA ASIGNATURA

#### TEMARIO TEÓRICO:

BLOQUE I Problemas de la recuperación de información en la Web y posibilidades de mejora en cuanto a la representación de recursos.

1. Recuperación de la Información en la Web: problemas y posibilidades de mejora.
2. Lenguajes de codificación descriptiva para la representación de la información.
  - o La edición electrónica. Los orígenes de SGML.
3. La norma ISO 8879, SGML (Standard Generalized Markup Language).



- o Gramática y sintaxis SGML.
- o Estructuras SGML.
  - Modelos de documento.
  - Definición del Tipo de Documento (DTD).

#### BLOQUE II XML.

1. La adaptación: XML (eXtensible Markup Language).
  - o Validación: DTDs y XML Schemas.
  - o El formato. La "convivencia": XML, HTML y CSS2.

#### BLOQUE III. Recuperación de Documentos Estructurados (XML)

1. La recuperación en los documentos estructurados.
2. Conceptos básicos de XPath.
3. Procesos.
  - o Indexación.
  - o Consulta.
  - o NEXI (*Narrowed Extended XPath*).
  - o Recuperación.

#### BLOQUE IV. "Familia" de especificaciones XML

1. Especificaciones complementarias
  - o El estilo y muchísimo más: XSL (eXtensible Stylesheet Language) y XSLT (eXtensible Stylesheet Language Transformation).
  - o La semántica (RDF-Resources Description Framework).

#### BLOQUE V La tecnología funcionando en conjunto

1. El proyecto Web Semántica
  - 1.1. Viabilidad
  - 1.2. Aplicaciones a servicios Web
2. Un Sistema Recomendación Semántico.
3. Conclusiones.

#### **TEMARIO PRÁCTICO:**

Coincidiendo con el contenido temático de la sesión, se realizarán las siguientes tareas prácticas (proyectos):

- Práctica 0: Desarrollo de un DTD SGML y un modelo de documentos correspondiente.
- Práctica 1: Diseño de nuestro primer XML y comprobación de *XML well formed*.
- Práctica 2: Diseño de un artículo científico en XML.
- Práctica 3: Mejoras de estilo con CSS2 y HTML
- Práctica 4: Expresiones NEXI para la consulta de documentos XML
- Práctica 5: Diseño y aplicación de plantilla XSLT
- Práctica 6: Ejercicio RDF
- Práctica 7 Diseño conceptual (con sus componentes RDF), de un Sistema de Recomendación Semántico

#### **BIBLIOGRAFÍA**

#### **BIBLIOGRAFÍA FUNDAMENTAL:**

- Berners-Lee, T. (1998 [B]). "Semantic Web Road map". <http://www.w3.org/DesignIssues/Semantic.html> (La "hoja de ruta" de la Web Semántica de Berners-Lee)
- Berners-Lee, T.; Hendler, J.; Lassila, O. (2001) "The Semantic Web: A new form of Web content that is meaningful to computers will unleash a revolution of new possibilities".



- Scientific American, May (Avance hacia la Web de datos)
- Feigenbaum, L.; Herman, I.; Hongsermeier, T.; Neumann, E.; Stephens, S. (2007). "The Semantic Web in Action". *Scientific American*, 297(6), pp. 90-97 (Desarrollos sobre la Web Semántica)
  - Fernández-Luna, J.M.; Huete-Guadix, J.F. (2001). "Recuperación XML". En: Cacheda, et. al. [eds]. *Recuperación de Información. Un Enfoque Práctico y Multidisciplinar*. Madrid: RA-MA, 415-453 (Trabajo fuente del bloque de recuperación estructurada).
  - Guha, R.; McCool, R.; Miller, E. (2003). "Semantic search". 12th International World Wide Web Conference (WWW2003), Budapest, Hungary 2003. pp. 700 - 709 (Recuperación de información en la Web Semántica).
  - Hendler, J. (2001). "Agents and the Semantic Web". *IEEE Intelligent Systems* 2001, marzo/abril, pp. 30-37 (El "padre" de la actuación de los agentes en la Web Semántica)
  - Hendler, J. (2002). "Ontologies on the Semantic Web". *IEEE Intelligent Systems* 2002, marzo-abril, pp.73-74 (El "centro" semántico de la Web Semántica, las Ontologías)
  - Herrera-Viedma, E.; Peis, E.; Olvera, M. D. et al. "Evaluating the Informative Quality of Web Sites by Fuzzy Computing with Words". En: *Lecture Notes in Artificial Intelligence*, 2003, 2663, pp.62-72 (Aplicación tecnologías Web Semántica y fuzzy logic).
  - Lassila, O.; Swick, R. (1999). "Resource Description Framework (RDF) Model and Syntax Specification: W3C Recommendation". <http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/> (Todo sobre RDF)
  - Peis, E.; Herrera-Viedma, E.; Morales-del-Castillo, J. M. (2007). Aproximación a la web semántica desde la perspectiva de la Documentación. *Investigación bibliotecológica*, 21(43), 47-71 (Muy útil para conocer la base tecnológica de la Web Semántica y discutir sobre su importancia para los profesionales de la información).
  - Peis, E.; Herrera-Viedma, E.; Hassan, Y.; Herrera, J. C. (2003). "Análisis de la web semántica: estado actual y requisitos futuros". *El Profesional de la Información*, 12(5), 368-376 (Análisis de las tecnologías básicas de la Web Semántica).
  - Peis, E.; Herrera-Viedma, E.; Hassan, Y.; Herrera, J. C. "Ontologías, Metadatos y Agentes: Recuperación Semántica de la Información". En: III Jornadas de Tratamiento y Recuperación de Información (JOTRI'2003). *Madrid, 8 y 9 de septiembre, 2003* (Análisis de la "capa lógica" de la Web Semántica)
  - Peis, E.; Morales-Del-Castillo, J.; Delgado-López, J. (2008). Semantic Recommender Systems - Analysis of the State of the Topic. *Hipertextnet*, 6, 1-5.
  - Peis, E.; Morales-del-Castillo, J.M.; Herrera-Viedma, E. (2011). "Tecnologías Web Semántica y Recuperación de Información". En: Cacheda, et. al. [eds]. *Recuperación de Información. Un Enfoque Práctico y Multidisciplinar*. Madrid: RA-MA, 641-673 (SPARQL y otras tecnologías para la recuperación en la Web Semántica)
  - Tejada-Lorente, A.; Porcel, C.; Peis, E.; Sanz, R. and Herrera-Viedma, Enrique. (2014). A quality based recommender system to disseminate information in a university digital library. *Information Sciences*, 261, pp. 52-69.

#### BIBLIOGRAFÍA COMPLEMENTARIA:

- Antinou, G.; van Harmelen, F. (2004). "A Semantic Web Primer". MIT Press (Monografía básica sobre Web Semántica)
- Bray, T.; Hollander, D.; Laymaná, A. (eds.) (1999). "Namespaces in XML". <http://www.w3.org/TR/REC-xml-names> (Aclaración sobre los Espacios de Nombre)
- Martín Quetglás, G.; Martín Benítez, I. (2005). "Curso de XML: Introducción al Lenguaje de la Web". Madrid: Pearson Alhambra (Monografía básica sobre XML).
- McGuinness, D. L.; van Harmelen, F. (2004). "OWL Web Ontology Language Overview. W3C Recommendation. 10 February 2004". <http://www.w3.org/TR/2004/REC-owl-features-20040210/> (El lenguaje para la construcción de ontologías en la Web Semántica)
- Shadbolt, N.; Hall, W.; Berners-Lee, T. (2006). "The Semantic Web Revisited". *IEEE Intelligent Systems*. May-June (Reformulación de los principios iniciales de la Web Semántica)



## ENLACES RECOMENDADOS (OPCIONAL)

- W3C-CSS2. (1998). "World Wide Web Consortium. Cascading Style Sheets, level 2 (CSS2). Recommendation 12 May 1998". <http://www.w3.org/TR/1998/REC-CSS2-19980512>
- W3C-DOM. (1998). "Document Object Model (DOM) Level 2 Specification, Version 1.0. Working Draft 28 December 1998". <http://www.w3.org/TR/1998/WD-DOM-Level-2-19981228>
- W3C-Metadata . (1998). "World Wide Web Consortium. Metadata and resource description. World Wide Web home page". <http://www.w3.org/Metadata>
- W3C-Namespaces. (1999). "World Wide Web Consortium. Namespaces in XML. Recommendation, 14 jan 1999". <http://www.w3.org/TR/1999/REC-xml-names-19990114/>
- W3C-RDF. "World Wide Web Consortium. Resource Description Framework. Proposed Recommendation". (2000). <http://www.w3.org/TR/PR-rdf-syntax/>
- W3C-Stylesheet. "World Wide Web Consortium. Associating stylesheets with XML documents. Working Draft". (2000). <http://www.w3.org/TR/WD-xml-stylesheet>
- W3C-XLink. (1999). "XML Linking Language (XLink). Working Draft 26 July 1999". <http://www.w3.org/1999/07/WD-xlink-19990726>
- W3C-XML. (1998). "World Wide Web Consortium. Extensible Markup Language (XML) 1.0 Recomendación 10 de febrero de 1998". <http://www.w3.org/TR/1998/REC-xml-19980210.html>
- W3C-XPath. (1999). "XML Path Language (XPath) Version 1.0 . <http://www.w3.org/TR/xpath/>

## METODOLOGÍA DOCENTE

El planteamiento metodológico será eminentemente práctico:

- Apoyándose en presentaciones, el profesor introducirá los temas del día, mediante lecciones y propondrá al estudiantado la resolución de tareas.
- Las tareas serán realizadas por el estudiantado.
- La viabilidad de la resolución de dichas tareas podrá ser comprobada por el propio estudiantado. En caso fallido, el estudiantado procederá a las oportunas correcciones y/o consultará al profesor). Una vez resuelta satisfactoriamente, el estudiantado "subirá" el producto final.
- El estudiantado interactuará con el resto de sus compañeros y con el profesor mediante el foro para expresar sus dudas, comentarios y/o compartir problemas en la resolución de tareas.
- El estudiantado también podrá comunicarse con el profesor mediante el correo interno.

Todas las propuestas citadas serán resueltas ayudándose de recursos bibliográficos, lecciones y tutoriales (tanto internos -aportados por el profesor-, como externos).

## EVALUACIÓN (INSTRUMENTOS DE EVALUACIÓN, CRITERIOS DE EVALUACIÓN Y PORCENTAJE SOBRE LA CALIFICACIÓN FINAL, ETC.)

### CONVOCATORIA ORDINARIA

Evaluación continua e interactiva de las tareas propuestas por el profesor.

Cada una de las propuestas tendrá una fecha aconsejada de carga en la plataforma y una fecha límite.

La calificación final vendrá determinada por la media ponderada de todas las tareas realizadas por el estudiantado.

### CONVOCATORIA EXTRAORDINARIA



La convocatoria extraordinaria se evaluará mediante el envío al profesor de la resolución de todas las tareas propuestas durante la impartición de la asignatura. Su calificación, conjunta, supondrá el 100% de la calificación final.

**DESCRIPCIÓN DE LAS PRUEBAS QUE FORMARÁN PARTE DE LA EVALUACIÓN ÚNICA FINAL ESTABLECIDA EN LA *NORMATIVA DE EVALUACIÓN Y DE CALIFICACIÓN DE LOS ESTUDIANTES DE LA UNIVERSIDAD DE GRANADA***

El profesor propondrá la realización de un proyecto, en el ordenador, en el que el estudiantado tenga que emplear la mayoría de las tecnologías estudiadas en la asignatura.

El estudiantado dispondrá de un tiempo limitado de 120 minutos para el desarrollo del proyecto.

Su calificación supondrá el 100% de la calificación final.

