



**DOCTORADO EN INGENIERÍA DEL LENGUAJE Y DEL CONOCIMIENTO**

**Área:** Ingeniería del Lenguaje y del Conocimiento

**Programa de Asignatura:** Técnicas y herramientas para el tratamiento de la información

**Código:** DILC10100

**Tipo:** Básica

**Créditos:** 9

**Fecha:** Enero 2016



### 1. DATOS GENERALES

<b>Nombre del Programa Educativo:</b>	Doctorado en Ingeniería del Lenguaje y del Conocimiento
<b>Modalidad Académica:</b>	Escolarizada
<b>Nombre de la Asignatura:</b>	Técnicas y herramientas para el tratamiento de la información
<b>Ubicación:</b>	Primer semestre (Obligatorio)

### 2. REVISIONES Y ACTUALIZACIONES

<b>Autores:</b>	José Arturo Olvera López Iván Olmos Pineda David Eduardo Pinto Avendaño
<b>Fecha de diseño:</b>	Enero 2016
<b>Fecha de la última actualización:</b>	
<b>Revisores:</b>	Darnes Vilariño Ayala Claudia Zepeda Cortés José Luis Carballido Carranza Maya Carrillo Ruiz Juan Manuel González Calleros Josefina Guerrero García Manuel Martín Ortiz Guillermo De Ita Luna
<b>Sinopsis de la revisión y/o actualización:</b>	



### 3. OBJETIVOS:

**Educacional:** Que el alumno conozca las técnicas para el tratamiento de información en texto, imágenes, audio y video, de tal manera que sea capaz de identificar las diferentes técnicas aplicables a cada tipo de datos para su limpieza, preparación y procesamiento.

**General:** Conocer de manera detallada los conceptos generales sobre la estructura de los datos en texto, imágenes, audio y video, para poder discernir entre las diferentes técnicas aplicables a cada uno de los formatos de datos, así como identificar espacios de oportunidad para el desarrollo de proyectos de investigación.

#### **Específicos:**

1. Conocer las características principales de los datos basados en texto, así como los métodos, técnicas y herramientas usadas para su manipulación
2. Conocer las características fundamentales de los datos basadas en audio, así como sus operaciones básicas
3. Conocer las características principales de las imágenes digitales, así como sus operaciones básicas
4. Conocer las características principales de los datos representados a través de video, así como sus operaciones básicas
5. Ser capaz de determinar líneas de acción para desarrollar investigación de punta en materia de ingeniería del lenguaje y del conocimiento, tomando como base los diferentes tipos de datos multimedia (texto, imágenes, audio y video).



#### 4. CONTENIDO

Unidad	Contenido Temático/Actividades de aprendizaje
1. Tratamiento de información textual	1.1. Introducción a la información textual 1.2. Información estructurada 1.3. Información semi-estructurada 1.3.1. SGML 1.3.2. Casos de estudio (XML, HTML, Latex, etc) 1.3.3. Herramientas para la manipulación de información semi-estructurada 1.3.3.1. Corpus Query Language (CQL) 1.3.3.2. Ontology Web Language (OWL) 1.3.3.3. Protegé 1.3.3.4. JQuery 1.4. Información no estructurada 1.4.1. Identificación de códigos de caracteres 1.4.2. Tokenizadores 1.4.3. Parsers 1.4.4. Otras herramientas para el tratamiento de información textual
2. Imágenes digitales	2.1. Introducción a las imágenes digitales: conceptos, modelos de color, representación 2.2. Formatos de las imágenes digitales 2.3. Operadores en imágenes digitales 2.3.1. Operadores puntuales 2.3.2. Operadores regionales 2.4. Herramientas para la manipulación de imágenes
3. Señales de audio	3.1. Introducción a las señales de audio 3.2. Formatos de audio 3.2.1. Con pérdida y sin pérdida 3.2.2. Con compresión y sin compresión 3.3. Extracción de información en señales de audio



**BENÉMERITA UNIVERSIDAD AUTÓNOMA DE PUEBLA**  
**FACULTAD DE CIENCIAS DE LA COMPUTACIÓN**

Unidad	Contenido Temático/Actividades de aprendizaje
	3.4. Herramientas para el procesamiento de audio
4. Video	4.1. Introducción al video: imágenes + sonido 4.2. Formatos de video 4.3. Operadores básicos en vídeo 4.4. Aplicaciones en vídeo 4.5. Herramientas para la manipulación de video

**Nota:** La bibliografía deberá ser amplia, actualizada (no mayor a cinco años) con ligas, portales y páginas de Internet, se recomienda usar los criterios del APA para referir la bibliografía.



<b>Bibliografía</b>	
<b>Básica</b>	<b>Complementaria</b>
<ul style="list-style-type: none"><li>• Jacob Perkins. Python Text Processing with NLTK 2.0 Cookbook: LITE Edition, 2011</li><li>• Gonzalez, R. C. Digital Image Processing (3rd Edition). Prentice Hall, 2007.</li><li>• Pratt, W. K. Introduction to Digital Image Processing. CRC press, 2013.</li><li>• Bovik A. C. The Essential Guide to Video Processing. Academic Press, 2009.</li><li>• Tekal A. M. Digital Video Processing. Prentice Hall, 1995.</li><li>• Woods J. W. Multidimensional Signal, Image, and Video Processing and Coding, Second Edition. Academic Press, 2011</li><li>• Boulanger, Richar; Lazzarini, Victor (Eds.). The Audio Programming Book. The MIT Press. 2010</li></ul>	<ul style="list-style-type: none"><li>• Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances, Wilson Wong, Wei Liu, M. Bennamoun, IGI-Global, 2011.</li><li>• Extensible Markup Language (XML) 1.1. Tim Bray, Jean Paoli, C. M. Sperberg-McQueen, Eve Maler, François Yergeau and John Cowan, eds. W3C Recommendation 16 August 2006, edited in place 29 September 2006.</li><li>• OWL 2 Web Ontology Language: XML Serialization. Boris Motik and Peter. F. Patel-Schneider, eds., 2008.</li><li>• OWL 2 Web Ontology Language: Direct Semantics. Boris Motik, Peter F. Patel-Schneider, and Bernardo Cuenca Grau , eds., 2008</li></ul>

## 5. CRITERIOS DE EVALUACIÓN

<b>Criterios</b>	<b>Porcentaje</b>
• Exámenes	15%
• Participación en clase	10%
• Tareas	20%
• Exposiciones	20%
• Trabajo de investigación y/o de intervención	20%
• Proyecto final	15%
<b>Total</b>	<b>100%</b>