

PROGRAMA DE CURSO

Código	Nombre			
CC72W	Compresión de Texto			
Nombre en Inglés				
Text Compression				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
	10	3		3
Requisitos			Carácter del Curso	
CC40A Diseño y Análisis de Algoritmos				
Resultados de Aprendizaje				
Entregar los fundamentos de teoría de la información relacionados a compresión de secuencias sin pérdida, y cubrir las familias principales de compresores de texto. Dar además una visión general de la indexación comprimida de texto.				

Metodología Docente	Evaluación General
Clases expositivas del profesor de cátedra, buscando la participación de los alumnos en pequeños problemas que se van proponiendo durante la exposición.	Individual, mediante la entrega de un informe de investigación y una tarea (implementación + evaluación de un método de compresión).

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Teoría de la Información	2
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Complejidad de Kolmogorov y sus limitaciones. Entropía, entropía relativa e información mutua. Teorema de Shannon. Desigualdad de Jensen y otras de teoría de la información. Propiedad de equipartición asintótica, entropía empírica. Procesos estocásticos, cadenas de Markov y su	Al término de esta unidad se espera que el alumno conozca los conceptos básicos de compresión, los que serán necesarios a lo largo del curso.	[1] Cap. 14, 2.1-2.7, 3, 4.1-4.2, 11.3, 13. [2] Cap. 2A

entropía. Procesos ergódicos. Compresión universal, convergencia y redundancia.		
---	--	--

Número	Nombre de la Unidad	Duración en Semanas
2	Compresión Estadística	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Modelamiento y codificación. Modelamiento estático, semiestático y dinámico. Códigos y sus propiedades. Desigualdad de Kraft. Códigos óptimos y redundancia. Códigos de Fano y de Shannon. Códigos de Huffman. Optimalidad. Codificación aritmética. Códigos dinámicos o adaptivos: Huffman y aritméticos. Modelamiento de lenguaje natural. Códigos multiarios. Códigos densos y su uso para bases de datos de texto. Compresión PPM.	Al término de esta unidad se espera que el alumno conozca los principales métodos de compresión estadística.	[1] Cap. 5.1-5.10, 13.3. [2] Cap. 2.2, 3, 5.1-5.4, 6

Número	Nombre de la Unidad	Duración en Semanas
3	Compresión Basada en Diccionarios	2.5
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Compresión por sustitución de substrings. Compresión Lempel-Ziv: LZ77, LZ78, LZW y variantes. Universalidad y convergencia. Compresión basada en gramáticas. Re-Pair, Sequitur. Compresión de texto estructurado.	Al término de esta unidad se espera que el alumno conozca los principales métodos de compresión basados en diccionarios.	[1] Cap. 13.4-13.5. [2] Cap. 8.

Número	Nombre de la Unidad	Duración en Semanas
4	Otros Métodos	1

Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Transformación de Burrows-Wheeler. Move-to-front y run-length encoding.	Al término de esta unidad se espera que el alumno conozca la transformación de Burrows-Wheeler, y los métodos de compresión asociados a ésta.	[4]

Número	Nombre de la Unidad	Duración en Semanas
5	Índices Comprimidos	5.5
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
Índices Invertidos: Estructura. Compresión de enteros. Direccionamiento a bloques y a documentos. Índices para el modelo vectorial y su compresión. Construcción. Arreglos de sufijos comprimidos: Arreglos de sufijos. Función Ψ . Wavelet trees y FM-Índices. LZ-índices. Compresión de árboles rotulados. Estructuras de datos comprimidas.	Al término de esta unidad se espera que el alumno conozca los conceptos básicos relacionados con la compresión de índices para texto.	[4]

Bibliografía
<p>[1] T. Cover and J. Thomas. Elements of Information Theory. Wiley, 2nd ed., 2006.</p> <p>[2] T. Bell, J. Cleary, and I. Witten. Text Compression. Prentice-Hall, 1990.</p> <p>[3] I. Witten, A. Moffat, and T. Bell. Managing Gigabytes. Morgan Kaufmann, 2nd ed., 1999.</p> <p>[4] G. Navarro and V. Mäkinen. Compressed full-text indexes. ACM Computing Surveys 39(1):article 2, 2007.</p> <p>[5] Artículos científicos.</p>

Vigencia desde:	
-----------------	--



Elaborado por: