

PROGRAMA DE CURSO

Código	Nombre			
CC7320	Estructuras de Datos Comprimidas			
Nombre en Inglés				
Compressed Data Structures				
SCT	Unidades Docentes	Horas de Cátedra	Horas Docencia Auxiliar	Horas de Trabajo Personal
6	10	3	0	7
Requisitos			Carácter del Curso	
CC4102 / Autor			Electivo	
Resultados de Aprendizaje				
<p>El curso entrega herramientas algorítmicas y de teoría de la información para el diseño de estructuras de datos que operan en espacio reducido, y presenta una variedad de estructuras compactas para representar diversos tipos de datos.</p> <p>El alumno que apruebe el curso habrá demostrado:</p> <ul style="list-style-type: none"> - Comprender la relevancia de las estructuras de datos comprimidas en el contexto de la memoria jerárquica o limitada. - Comprender las técnicas más relevantes para el diseño de estructuras de datos comprimidas. - Conocer el funcionamiento de un conjunto relevante de estructuras de datos comprimidas existentes para varios problemas fundamentales. - Comprender las diferencias entre la teoría y la práctica en este ámbito. 				

Metodología Docente	Evaluación General
Clases expositivas del profesor de cátedra y de los mismos alumnos, éstos con tutoría del profesor, que velará por que los contenidos se presenten adecuadamente, tomando parte de la clase a su cargo de ser necesario.	Se evalúan las presentaciones de los alumnos con 50% y otro 50% con un trabajo final individual, donde se estudia una determinada estructura, se la presenta en forma conceptual y teórica, y se la implementa y compara contra el estado del arte o contra su contraparte clásica.

Unidades Temáticas

Número	Nombre de la Unidad	Duración en Semanas
1	Conceptos Básicos	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Relevancia de las estructuras de datos comprimidas en el contexto actual. 2. Concepto de estructura de datos compacta. Ejemplos de casos exitosos. 3. Concepto de entropía de peor caso, entropía empírica de orden cero y orden superior, y otros. Ejemplos. 4. Métodos de compresión clásicos: Huffman, LZ78, Re-Pair, codificación de diferencias.	Comprender la relevancia de las estructuras de datos comprimidas. Conocer algunos ejemplos existentes. Comprender cuándo la compresión es posible y por qué razón. Conocer algunos métodos clásicos de compresión.	[1]

Número	Nombre de la Unidad	Duración en Semanas
2	Bitmaps, Secuencias y Permutaciones	4,5
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Rank y select en secuencias binarias. Aplicaciones. 2. Soluciones de tiempo constante usando $n+o(n)$ bits. 3. Soluciones usando espacio comprimido, constantes y no constantes. 4. Rank y select sobre secuencias de símbolos. Aplicaciones.	Conocer las principales representaciones de secuencias binarias y generales, y derivados como permutaciones, grillas y relaciones binarias.	[1]

<p>5. El wavelet tree. Variantes.</p> <p>6. Aplicaciones del wavelet tree: grillas, relaciones binarias, grafos.</p> <p>7. Permutaciones.</p> <p>8. Representación de secuencias de Golynski et al.</p> <p>9. Particionamiento del alfabeto.</p> <p>10. Soportando dinamismo (inserción y borrado).</p> <p>11. Acceso directo a representaciones de largo variable.</p>		
---	--	--

Número	Nombre de la Unidad	Duración en Semanas
3	Arboles	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
<p>1. Operaciones deseables. Por qué son compresibles.</p> <p>2. Representación de paréntesis balanceados. Traducción de operaciones. Solución de tiempo constante.</p> <p>3. Representación DFUDS.</p> <p>4. Representación LOUDS.</p> <p>5. Representación con MinMax trees.</p> <p>6. Arboles rotulados.</p>	<p>Conocer las diversas representaciones de árboles generales y su funcionamiento.</p>	<p>[1]</p>

Número	Nombre de la Unidad	Duración en Semanas
4	Grafos y grillas	1
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Representaciones clásicas y operaciones a soportar. 2. Representación basada en secuencias. 3. Representación basada en k2-trees. 4. Extensión para representar puntos en grillas.	Conocer algunas de las representaciones existentes para grafos generales, dirigidos y no dirigidos. Conocer la forma en que se extienden para representar grillas discretas de puntos.	[1]

Número	Nombre de la Unidad	Duración en Semanas
5	Colecciones de Texto	3
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. El problema. Arreglos de sufijos. 2. La función Psi y su compresión. El Compressed Suffix Array. 3. La Transformación de Burrows-Wheeler. 4. Búsqueda reversa y el FM-index. Variantes. 5. El LZ-Index. 6. Colecciones dinámicas (insertar y borrar documentos).	Comprender el problema de manipular una colección de texto general. Conocer las principales soluciones comprimidas.	[1]

Número	Nombre de la Unidad	Duración en Semanas
6	Arboles de Sufijos	2.5
Contenidos	Resultados de Aprendizajes de la Unidad	Referencias a la Bibliografía
1. Arboles de sufijos generales. 2. El problema del Lowest Common Ancestor (LCA) y estructuras para Range Minimum Queries (RMQ). 3. Representación de paréntesis. 4. Representación basada en intervalos. 5. Representación basada en sampling.	Comprender la estructura de árbol de sufijo, su utilidad, y las representaciones comprimidas existentes.	[1]

Bibliografía
[1] G. Navarro. Compact Data Structures – A Practical Approach. Cambridge, 2016.

Vigencia desde:	Otoño 2018
Elaborado por:	Gonzalo Navarro