

## 1. CURSO

CS312. Estructuras de Datos Avanzadas (Obligatorio)

## 2. INFORMACIÓN GENERAL

2.1 Curso	:	CS312. Estructuras de Datos Avanzadas
2.2 Semestre	:	6 <sup>to</sup> Semestre.
2.3 Créditos	:	4
2.4 horas	:	2 HT; 4 HP;
2.5 Duración del periodo	:	16 semanas
2.6 Condición	:	Obligatorio
2.7 Modalidad de aprendizaje	:	Híbrido
2.8 Prerrequisitos	:	CS212. Análisis y Diseño de Algoritmos. (5 <sup>to</sup> Sem) CS212. Análisis y Diseño de Algoritmos. (5 <sup>to</sup> Sem)

## 3. PROFESORES

Atención previa coordinación con el profesor

## 4. INTRODUCCIÓN AL CURSO

Los algoritmos y estructuras de datos son una parte fundamental de la ciencia de la computación que nos permiten organizar la información de una manera más eficiente, por lo que es importante para todo profesional del área tener una sólida formación en este aspecto.

En el curso de estructuras de datos avanzadas nuestro objetivo es que el alumno conozca y analice estructuras complejas, como los Métodos de Acceso Multidimensional, Métodos de Acceso Espacio-Temporal y Métodos de Acceso Métrico, etc.

## 5. OBJETIVOS

- Que el alumno entienda, diseñe, implemente, aplique y proponga estructuras de datos innovadoras para solucionar problemas relacionados al tratamiento de datos multidimensionales, recuperación de información por similitud, motores de búsqueda y otros problemas computacionales.

## 6. RESULTADOS DEL ESTUDIANTE

- 1) S.O. Analizar un problema computacional complejo y aplicar los principios computacionales y otras disciplinas relevantes para identificar soluciones. (**Usar**)
- 2) S.O. Diseñar, implementar y evaluar una solución basada en computación para cumplir con un conjunto determinado de requisitos computacionales en el contexto de las disciplinas del programa. (**Familiarizarse**)
- 6) S.O. Aplicar la teoría de la computación y los fundamentos del desarrollo de software para producir soluciones basadas en computación. (**Familiarizarse**)

## 7. TEMAS

<b>Unidad 1: Multidimensional Data (16)</b>	
<b>Resultados esperados: 1,2</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Introducción al curso.</li> <li>• Introducción a datos multidimensionales.</li> <li>• Maldición de la dimensionalidad.</li> </ul>	<ul style="list-style-type: none"> <li>• Introducir la trascendencia de la representación multidimensional de datos. [Usar]</li> <li>• Entender la complejidad de lidiar con datos multidimensional y de alta dimensión.[Usar]</li> <li>• Entender la maldición de la dimensionalidad, y su impacto en el indizado de grandes volúmenes de datos.[Usar]</li> <li>• Presentar y discutir aplicaciones reales de datos multidimensionales en motores de búsqueda.[Usar]</li> </ul>
<b>Lecturas :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unidad 2: Multidimensional Acces Data Structures (16)</b>	
<b>Resultados esperados: 1,2</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Introducción a estructuras de datos espaciales.</li> <li>• Estructuras espaciales, Quadtree,Octree y visualización.</li> <li>• Kd-Tree.</li> <li>• Introducción a R-Tress.</li> <li>• R tree (Guttman).</li> <li>• R+ tree.</li> <li>• R* tree.</li> <li>• Variación R*-tree y relación con paginación y tamaño de bloques.</li> <li>• X-tree.</li> </ul>	<ul style="list-style-type: none"> <li>• Introducir los fundamentos teóricos de estructuras de datos espaciales.</li> <li>• Entender los beneficios y limitaciones deestructuras de datos espaciales basadas en árbol.</li> <li>• Implementar diferentes estructuras de datos para el indizado de grandes volúmenes de datos.</li> <li>• Entender los fundamentos e implementar estrategias de búsqueda como vecinos mas próximos y búsquedas por rango.</li> </ul>
<b>Lecturas :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unidad 3: Approximate Access Methods (20)</b>	
<b>Resultados esperados: 1,2</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Métodos de Acceso Métrico para distancias discretas</li> <li>• Métodos de Acceso Métrico para distancias continuas</li> </ul>	<ul style="list-style-type: none"> <li>• Que el alumno entienda conozca e implemente algunos métodos de acceso métrico[Usar]</li> <li>• Que el alumno entienda la importancia de estos Métodos de Acceso para la Recuperación de Información por Similitud[Usar]</li> </ul>
<b>Lecturas :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

Unidad 4: Métodos de Acceso Aproximados (20)	
Resultados esperados: 1,2	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> <li>• Space Filling Curves: Hilbert curve y Z-order</li> <li>• Proyecciones y complejidad.</li> <li>• Locally sensitive hashing (LSH)</li> </ul>	<ul style="list-style-type: none"> <li>• Entender, conocer e implementar algunos métodos de acceso aproximados.</li> <li>• Entender la importancia de estos métodos de Acceso para la recuperación de información por similitud en entornos donde la escalabilidad sea una factor muy importante.</li> </ul>
<b>Lecturas :</b> [Sam06], [PI06], [Zez+07]	

Unidad 5: Clustering (8)	
Resultados esperados: 1,2	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> <li>• Introducción a Clustering.</li> <li>• Kmeans y DBScan.</li> <li>• Clustering Applications.</li> <li>• Clustering Ensemble.</li> </ul>	<ul style="list-style-type: none"> <li>• Introducir los fundamentos teóricos para el clustering de datos multidimensionales.</li> <li>• Implementar diferentes estrategias para el clustering de datos multidimensionales, como basados en partición, en jerarquía o en densidad.</li> <li>• Entender los fundamentos, aplicaciones e implementar ensambles de métodos de clustering.</li> <li>• Implementar ensambles de métodos de clustering con datos reales.</li> </ul>
<b>Lecturas :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

Unidad 6: Temporal Data Structures (8)	
Resultados esperados: 1,2	
Temas	Objetivos de Aprendizaje
<ul style="list-style-type: none"> <li>• Introducción a Estructuras de datos temporales.</li> <li>• Versionando la estructura de Datos.</li> <li>• Persistencia</li> <li>• Retroactividad</li> </ul>	<ul style="list-style-type: none"> <li>• Introducir los fundamentos teóricos de estructuras de datos temporales.</li> <li>• Entender, discutir e implementar Persistencia y sus tipos.</li> <li>• Entender, discutir e implementar Retroactividad y sus tipos.</li> <li>• Entender y discutir los beneficios y limitaciones entre persistencia y retroactividad.</li> </ul>
<b>Lecturas :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unidad 7: Final Talks (8)</b>	
<b>Resultados esperados: 1,2</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje</b>
<ul style="list-style-type: none"> <li>• Seminarios de trabajo de investigación.</li> </ul>	<ul style="list-style-type: none"> <li>• Investigar sobre nuevos métodos para el indizado de grandes volúmenes de datos complejos.</li> <li>• Presentar y dirigir la discusión sobre métodos para indizados de Big Data investigado.</li> </ul>
<b>Lecturas :</b> [SCuadros2004Implementing], [Knu07a], [Knu07b], [Gam+94]	

## 8. PLAN DE TRABAJO

### 8.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

### 8.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

### 8.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

## 9. SISTEMA DE EVALUACIÓN

\*\*\*\*\* EVALUATION MISSING \*\*\*\*\*

## 10. BIBLIOGRAFÍA BÁSICA

- [Cua+04] Ernesto Cuadros-Vargas et al. "Implementing data structures: An incremental approach". <http://socios.spc.org.pe/ecuadros/cursos/pdfs/>. 2004.
- [Gam+94] Erich Gamma et al. *Design Patterns: Elements of Reusable Object-Oriented Software*. Computing Series. ISBN-10: 0201633612. Addison-Wesley Professional, Nov. 1994.
- [Knu07a] Donald Ervin Knuth. *The Art of Computer Programming, Fundamental Algorithms*. 3rd. Vol. I. 0-201-89683-4. Addison-Wesley, Feb. 2007.
- [Knu07b] Donald Ervin Knuth. *The Art of Computer Programming, Sorting and Searching*. 2nd. Vol. II. 0-201-89685-0. Addison-Wesley, Feb. 2007.
- [PI06] Trevor Darrell PGregory Shakhnarovich and Piotr Indyk. *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice*. 1st. ISBN 0-262-19547-X. MIT Press, Mar. 2006.
- [Sam06] Hanan Samet. *Foundations of Multidimensional and Metric Data Structures*. Illustrated. Elsevier/Morgan Kaufmann, Aug. 2006. ISBN: 9780123694461. URL: <http://books.google.com.pe/books?id=v0-NRRKHG84C>.
- [Zez+07] Pavel Zezula et al. *Similarity Search: The Metric Space Approach*. 1st. ISBN-10: 0387291466. Springer, Nov. 2007.