



## National University of Engineering (UNI)

School of Artificial Intelligence

Syllabus 2024-I

### 1. COURSE

CS312. Advanced Data Structures (Mandatory)

### 2. GENERAL INFORMATION

2.1 Course	:	CS312. Advanced Data Structures
2.2 Semester	:	6 <sup>th</sup> Semester.
2.3 Credits	:	4
2.4 Horas	:	2 HT; 4 HP;
2.5 Duration of the period	:	16 weeks
2.6 Type of course	:	Mandatory
2.7 Learning modality	:	Face to face
2.8 Prerequisites	:	CS212. Analysis and Design of Algorithms. (5 <sup>th</sup> Sem)

### 3. PROFESSORS

Meetings after coordination with the professor

### 4. INTRODUCTION TO THE COURSE

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. GOALS

- 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. COMPETENCES

- 1) Analizar un problema computacional complejo y aplicar los principios computacionales y otras disciplinas relevantes para identificar soluciones. (**Usar**)
- 2) 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) Aplicar la teoría de la computación y los fundamentos del desarrollo de software para producir soluciones basadas en computación. (**Familiarizarse**)

### 7. TOPICS

<b>Unit 1: Multidimensional Data (16 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</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>Readings :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unit 2: Multidimensional Acces Data Structures (16 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</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>Readings :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unit 3: Approximate Access Methods (20 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</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>Readings :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unit 4: Métodos de Acceso Aproximados (20 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</b>
<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>Readings :</b> [Sam06], [PI06], [Zez+07]	

<b>Unit 5: Clustering (8 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</b>
<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>Readings :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unit 6: Temporal Data Structures (8 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</b>
<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>Readings :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

<b>Unit 7: Final Talks (8 hours)</b>	
<b>Competences Expected: 1,2</b>	
<b>Topics</b>	<b>Learning Outcomes</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>Readings :</b> [Cua+04], [Knu07a], [Knu07b], [Gam+94]	

## 8. WORKPLAN

### 8.1 Methodology

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

### 8.2 Theory Sessions

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

### 8.3 Practical Sessions

The practical sessions are held in class where a series of exercises and/or practical concepts are developed through problem solving, problem solving, specific exercises and/or in application contexts.

## 9. EVALUATION SYSTEM

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

## 10. BASIC BIBLIOGRAPHY

- [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.