



Universidad de Ingeniería y Tecnología
Escuela Profesional de
Ciencia de la Computación
Silabo del curso – Periodo Académico 2018-I

1. **Código del curso y nombre:** CB309. Bioinformática y Bioestadística
2. **Créditos:** 4
3. **Horas de Teoría y Laboratorio:** 2 HT; 4 HP;
4. **Docente(s)**

Atención previa coordinación con el profesor

5. Bibliografía

- [Alu06] Srinivas Aluru, ed. *Handbook of Computational Molecular Biology*. Computer and Information Science Series. Boca Raton, FL: Chapman & Hall, CRC, 2006.
- [CB00] P. Clote and R. Backofen. *Computational Molecular Biology: An Introduction*. 279 pages. John Wiley & Sons Ltd., 2000.
- [Dur+98] R. Durbin et al. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998, p. 357. ISBN: 9780521629713.
- [Kro+94] Anders Krogh et al. "Hidden Markov Models in Computational Biology, Applications to Protein Modeling". In: *J Molecular Biology* 235 (1994), pp. 1501–1531.
- [Pev00] Pavel A. Pevzner. *Computational Molecular Biology: an Algorithmic Approach*. Cambridge, Massachusetts: The MIT Press, 2000.
- [SM97] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. Boston: PWS Publishing Company, 1997, pp. I–XIII, 1–296. ISBN: 978-0-534-95262-4.

6. Información del curso

- (a) **Breve descripción del curso** El uso de métodos computacionales en las ciencias biológicas se ha convertido en una de las herramientas claves para el campo de la biología molecular, siendo parte fundamental en las investigaciones de esta área.

En Biología Molecular, existen diversas aplicaciones que involucran tanto al ADN, al análisis de proteínas o al secuenciamiento del genoma humano, que dependen de métodos computacionales. Muchos de estos problemas son realmente complejos y tratan con grandes conjuntos de datos.

Este curso puede ser aprovechado para ver casos de uso concretos de varias áreas de conocimiento de Ciencia de la Computación como: Lenguajes de Programación (PL), Algoritmos y Complejidad (AL), Probabilidades y Estadística, Manejo de Información (IM), Sistemas Inteligentes (IS).

- (b) **Prerrequisitos:** CS212. Análisis y Diseño de Algoritmos. (5^{to} Sem)

- (c) **Tipo de Curso:** Obligatorio

7. Competencias

- Que el alumno tenga un conocimiento sólido de los problemas biológicos moleculares que desafían a la computación.
- Que el alumno sea capaz de abstraer la esencia de los diversos problemas biológicos para plantear soluciones usando sus conocimientos de Ciencia de la Computación

8. Contribución a los resultados (Outcomes)

- a) Aplicar conocimientos de computación y de matemáticas apropiadas para la disciplina. (**Usar**)

b) Analizar problemas e identificar y definir los requerimientos computacionales apropiados para su solución. (**Evaluar**)

l) Desarrollar principios investigación en el área de computación con niveles de competitividad internacional. (**Evaluar**)

9. Competencias (IEEE)

C1. La comprensión intelectual y la capacidad de aplicar las bases matemáticas y la teoría de la informática (*Computer Science*).⇒ **Outcome a,b**

C3. Una comprensión intelectual de, y el aprecio por el papel central de los algoritmos y estructuras de datos.⇒ **Outcome b,l**

C5. Capacidad para implementar algoritmos y estructuras de datos en el software.⇒ **Outcome a,b**

10. Lista de temas a estudiar en el curso

1. Introducción a la Biología Molecular
2. Comparación de Secuencias
3. Árboles Filogenéticos
4. Ensamblaje de Secuencias de ADN
5. Estructuras secundarias y terciarias
6. Modelos Probabilísticos en Biología Molecular

11. Metodología y Evaluación

Metodología:

Sesiones Teóricas:

El desarrollo de las sesiones teóricas está focalizado en el estudiante, a través de su participación activa, resolviendo problemas relacionados al curso con los aportes individuales y discutiendo casos reales de la industria. Los alumnos desarrollarán a lo largo del curso un proyecto de aplicación de las herramientas recibidas en una empresa.

Sesiones de Laboratorio:

Las sesiones prácticas se desarrollan en laboratorio. Las prácticas de laboratorio se realizan en equipos para fortalecer su comunicación. Al inicio de cada laboratorio se explica el desarrollo de la práctica y al término se destaca las principales conclusiones de la actividad en forma grupal.

Exposiciones individuales o grupales:

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.

Lecturas:

A lo largo del curso se proporcionan diferentes lecturas, las cuales son evaluadas. El promedio de las notas de las lecturas es considerado como la nota de una práctica calificada. El uso del campus virtual UTEC Online permite a cada estudiante acceder a la información del curso, e interactuar fuera de aula con el profesor y con los otros estudiantes.

Sistema de Evaluación:

12. Contenido

Unidad 1: Introducción a la Biología Molecular (4)	
Competencias esperadas: CS1	
Objetivos de Aprendizaje	Tópicos
<ul style="list-style-type: none"> • Lograr un conocimiento general de los tópicos más importantes en Biología Molecular. [Familiarizarse] • Entender que los problemas biológicos son un desafío al mundo computacional. [Evaluar] 	<ul style="list-style-type: none"> • Revisión de la química orgánica: moléculas y macromoléculas, azúcares, ácidos nucleicos, nucleótidos, ARN, ADN, proteínas, aminoácidos y niveles de estructura en las proteínas. • El dogma de la vida: del ADN a las proteínas, transcripción, traducción, síntesis de proteínas • Estudio del genoma: Mapas y secuencias, técnicas específicas
Lecturas : [CB00], [SM97]	

Unidad 2: Comparación de Secuencias (4)	
Competencias esperadas: CS2	
Objetivos de Aprendizaje	Tópicos
<ul style="list-style-type: none"> • Entender y solucionar el problema de alineamiento de un par de secuencias. [Usar] • Comprender y solucionar el problema de alineamiento de múltiples secuencias. [Usar] • Conocer los diversos algoritmos de alineamiento de secuencias existentes en la literatura. [Familiarizarse] 	<ul style="list-style-type: none"> • Secuencias de nucleótidos y secuencias de aminoácidos. • Alineamiento de secuencias, el problema de alineamiento por pares, búsqueda exhaustiva, Programación dinámica, alineamiento global, alineamiento local, penalización por gaps • Comparación de múltiples secuencias: suma de pares, análisis de complejidad por programación dinámica, heurísticas de alineamiento, algoritmo estrella, algoritmos de alineamiento progresivo.
Lecturas : [CB00], [SM97], [Pev00]	

Unidad 3: Árboles Filogenéticos (4)	
Competencias esperadas: CS2	
Objetivos de Aprendizaje	Tópicos
<ul style="list-style-type: none"> • Comprender el concepto de filogenia, árboles filogenéticos y la diferencia metodológica entre biología y biología molecular. [Familiarizarse] • Comprender el problema de reconstrucción de árboles filogenéticos, conocer y aplicar los principales algoritmos para reconstrucción de árboles filogenéticos. [Evaluar] 	<ul style="list-style-type: none"> • Filogenia: Introducción y relaciones filogenéticas. • Árboles Filogenéticos: definición, tipo de árboles, problema de búsqueda y reconstrucción de árboles • Métodos de Reconstrucción: métodos por parsimonia, métodos por distancia, métodos por máxima verosimilitud, confianza de los árboles reconstruidos
Lecturas : [CB00], [SM97], [Pev00]	

Unidad 4: Ensamblaje de Secuencias de ADN (4)	
Competences esperadas: CS2	
Objetivos de Aprendizaje	Tópicos
<ul style="list-style-type: none"> • Comprender el desafío computacional que ofrece el problema de Ensamblaje de Secuencias. [Familiarizarse] • Entender el principio de modelo formal para ensamblaje. [Evaluar] • Conocer las principales heurísticas para el problema de ensamblaje de secuencias ADN [Usar] 	<ul style="list-style-type: none"> • Fundamento biológico: caso ideal, dificultades, métodos alternativos para secuenciación de ADN • Modelos formales de ensamblaje: <i>Shortest Common Superstring, Reconstruction, Multicontig</i> • Algoritmos para ensamblaje de secuencias: representación de overlaps, caminos para crear <i>superstrings</i>, algoritmo voraz, grafos acíclicos. • Heurísticas para ensamblaje: búsqueda de sobreposiciones, ordenación de fragmentos, alineamientos y consenso.
Lecturas : [SM97], [Alu06]	

Unidad 5: Estructuras secundarias y terciarias (4)	
Competences esperadas: CS2	
Objetivos de Aprendizaje	Tópicos
<ul style="list-style-type: none"> • Conocer las estructuras protéicas y la necesidad de métodos computacionales para la predicción de la geometría. [Familiarizarse] • Conocer los algoritmos de solución de problemas de predicción de estructuras secundarias ARN, y de estructuras en proteínas. [Evaluar] 	<ul style="list-style-type: none"> • Estructuras moleculares: primaria, secundaria, terciaria, cuaternaria. • Predicción de estructuras secundarias de ARN: modelo formal, energía de pares, estructuras con bases independientes, solución con Programación Dinámica, estructuras con bucles. • <i>Protein folding</i>: Estructuras en proteínas, problema de <i>protein folding</i>. • <i>Protein Threading</i>: Definiciones, Algoritmo <i>Branch & Bound</i>, <i>Branch & Bound</i> para <i>protein threading</i>. • <i>Structural Alignment</i>: definiciones, algoritmo DALI
Lecturas : [SM97], [CB00], [Alu06]	

Unidad 6: Modelos Probabilísticos en Biología Molecular (4)	
Competences esperadas: CS2	
Objetivos de Aprendizaje	Tópicos
<ul style="list-style-type: none"> • Revisar conceptos de Modelos Probabilísticos y comprender su importancia en Biología Molecular Computacional. [Evaluar] • Conocer y aplicar Modelos Ocultos de Markov para varios análisis en Biología Molecular. [Usar] • Conocer la aplicación de modelos probabilísticos en Filogenia y compararlos con modelos no probabilísticos [Evaluar] 	<ul style="list-style-type: none"> • Probabilidad: Variables aleatorias, Cadenas de Markov, Algoritmo de Metropoli-Hasting, Campos Aleatorios de Markov y Muestreador de Gibbs, Máxima Verosimilitud. • Modelos Ocultos de Markov (HMM), estimación de parámetros, algoritmo de Viterbi y método Baum-Welch, Aplicación en alineamientos de pares y múltiples, en detección de Motifs en proteínas, en ADN eucariótico, en familias de secuencias. • Filogenia Probabilística: Modelos probabilísticos de evolución, verosimilitud de alineamientos, verosimilitud para inferencia, comparación de métodos probabilísticos y no probabilísticos
Lecturas : [Dur+98], [CB00], [Alu06], [Kro+94]	