



Sociedad Peruana de Computación (SPC)

Programa Profesional de
Ciencia de la Computación
Sílabo 2021-I

1. CURSO

CB309. Bioinformática (Obligatorio)

2. INFORMACIÓN GENERAL

- 2.1 Créditos : 2
- 2.2 Horas de teoría : 1 (Semanal)
- 2.3 Horas de práctica : 2 (Semanal)
- 2.4 Duración del periodo : 16 semanas
- 2.5 Condición : Obligatorio
- 2.6 Modalidad : Presencial
- 2.7 Prerrequisitos :
 - CS212. Análisis y Diseño de Algoritmos. (5^{to} Sem)
 - MA307. Matemática aplicada a la computación. (6^{to} Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL 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).

5. OBJETIVOS

- 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

6. COMPETENCIAS

- a) Aplicar conocimientos de computación y de matemáticas apropiadas para la disciplina. (**Evaluar**)
- b) Analizar problemas e identificar y definir los requerimientos computacionales apropiados para su solución. (**Evaluar**)
- h) Incorporarse a un proceso de aprendizaje profesional continuo. (**Usar**)
- i) Utilizar técnicas y herramientas actuales necesarias para la práctica de la computación. (**Usar**)
- j) Aplicar la base matemática, principios de algoritmos y la teoría de la CS en el modelamiento y diseño de sistemas. (**Usar**)

7. COMPETENCIAS ESPECÍFICAS

- a58) Aplicar estructuras de datos y algoritmos para tratar problemas biológicos desde la perspectiva computacional.
- b4) Identificar y aplicar de forma eficiente diversas estrategias algorítmicas y estructuras de datos para la solución de un problema dadas ciertas restricciones de espacio y tiempo.

h2) Demostrar que tiene capacidad de aprender a aprender de forma autónoma.

i15) Aplicar programación dinámica, heurísticas y métodos de detección de agrupamientos (*clustering*).

j2) Aplicar teoría de grafos y árboles para la optimización y resolución de problemas.

8. TEMAS

Unidad 1: Introducción a la Biología Molecular (4)	
Competencias esperadas: CS1	
Temas	Objetivos de Aprendizaje
<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	<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]
Lecturas : [CB00], [SM97]	

Unidad 2: Comparación de Secuencias (4)	
Competencias esperadas: CS2	
Temas	Objetivos de Aprendizaje
<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.	<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]
Lecturas : [CB00], [SM97], [Pev00]	

Unidad 3: Árboles Filogenéticos (4)	
Competencias esperadas: CS2	
Temas	Objetivos de Aprendizaje
<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	<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]
Lecturas : [CB00], [SM97], [Pev00]	

Unidad 4: Ensamblaje de Secuencias de ADN (4)	
Competencias esperadas: CS2	
Temas	Objetivos de Aprendizaje
<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. 	<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]
Lecturas : [SM97], [Alu06]	

Unidad 5: Estructuras secundarias y terciarias (4)	
Competencias esperadas: CS2	
Temas	Objetivos de Aprendizaje
<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, Branch & Bound</i> para <i>protein threading</i>. • <i>Structural Alignment</i>: definiciones, algoritmo DALI 	<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]
Lecturas : [SM97], [CB00], [Alu06]	

Unidad 6: Modelos Probabilísticos en Biología Molecular (4)	
Competencias esperadas: CS2	
Temas	Objetivos de Aprendizaje
<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 Baul-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 	<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 comparalos con modelos no probabilísticos [Evaluar]
Lecturas : [Dur+98], [CB00], [Alu06], [Kro+94]	

9. PLAN DE TRABAJO

9.1 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.

9.2 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.

9.3 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.

10. SISTEMA DE EVALUACIÓN

***** EVALUATION MISSING *****

11. BIBLIOGRAFÍA BÁSICA

- [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.