

QUILMES, 23 FEB 2010

VISTO el Expediente N° 827-1474/09, y

**CONSIDERANDO:**

Que por el citado Expediente se tramita la aprobación del Curso de Doctorado denominado "Algoritmos en Inmunoinformática".

Que a través de la Resolución (CS) N° 283/05, se aprueba el Reglamento de Cursos y Seminarios de Posgrado de la Universidad.

Que el mencionado curso constituye un aporte relevante a la formación de posgrado en las especialidades involucradas.

Que los antecedentes académicos y profesionales de los docentes a cargo del dictado del mismo, garantizan calidad y solvencia en el desarrollo de los contenidos especificados.

Que la evaluación del citado curso ha cumplido con los requisitos estipulados en el Art. 6° del Reglamento de Cursos y Seminarios de Posgrado de esta Casa de Altos Estudios.

Que la presente se dicta en virtud de las atribuciones conferidas al Rector por el Art. 72° del Estatuto Universitario.

Por ello,

**EL RECTOR DE LA UNIVERSIDAD NACIONAL DE QUILMES**

**RESUELVE:**

ARTICULO 1°: Aprobar el dictado del Curso de Doctorado denominado "Algoritmos en Inmunoinformática" cuyo programa y características generales se detallan en el Anexo de la presente Resolución.

ARTICULO 2°: Designar como docente expositor para el dictado del curso al Dr. Morten Nielsen y como docente coordinador al Dr. Gustavo Parisi.

ARTICULO 3°: Disponer que el curso tendrá una duración total de 64 (sesenta y cuatro) horas y que se podrá dictar hasta el ciclo lectivo 2012.

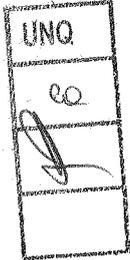


—00108

ARTICULO 4º: Establecer un cupo máximo de 20 alumnos. En el caso que los postulantes excedan esa cifra, el docente a cargo realizará la selección correspondiente.

ARTICULO 5º: Regístrese, practíquense las comunicaciones de estilo y archívese.

RESOLUCION (R) Nº: 00108



Mg. Alfredo Alfonso  
Secretario General  
UNIVERSIDAD NACIONAL DE QUILMES



Gustavo Eduardo Lugones  
Rector  
Universidad Nacional de Quilmes

**ANEXO**

**Título del Curso de Doctorado:** "Algoritmos en Inmunoinformática"

**Lugar de Realización:** Universidad Nacional de Quilmes - Roque Sáenz Peña 352, Bernal.

**Docente Expositor:** Dr. Morten Nielsen.

**Docente coordinador:** Dr. Gustavo Parisi.

**Carga horaria:** 64 hs.

**Fecha de realización:** año 2010 con aprobación hasta el 2012.

**Destinatarios:** Biotecnólogos, Bioquímicos, Biólogos moleculares y carreras afines.

**Objetivos:**

- Entender los detalles de los algoritmos utilizados en la bioinformática aplicada a la inmunología.
- Desarrollar aplicaciones o programas que utilizan los algoritmos vistos en el curso.
- Identificar el tipo de algoritmo más adecuado para describir un problema biológico determinado.
- Comprender los conceptos de redundancia de datos y la reducción de la homología.
- Desarrollar algoritmos de predicción para resolver un problema biológico determinado.
- Aplicación y desarrollo de herramientas de predicción en un nivel de detalle adecuado utilizando los siguientes algoritmos: Programación dinámica (Dynamic programming), la agrupación de secuencias (Sequence clustering), las matrices de peso (Weight matrices), redes neuronales artificiales [Artificial neural networks), modelos ocultos de Markov (Hidden Markov models),
- Diseño de un proyecto en el que un problema biológico se analiza utilizando uno o varios algoritmos de aprendizaje automático.
- Implementar, documentar y presentar el proyecto de curso.



## Bibliografía y Contenidos:

### Contenidos generales

- Programación dinámica (Dynamic programming): Needleman-Wunsch, Smith-Waterman, and alignment heuristics
- La redundancia de datos y la reducción de la homología (Data redundancy and homology reduction): Hobohm and other clustering algorithms
- Matrices Peso (Weight matrices): Sequence weighting, pseudo count correction for low counts, Gibbs sampling, and Psi-Blast
- Modelos ocultos de Harkov (Hidden Markov Models): Model construction, Viterbi decoding, and posterior decoding, and Baum Welsh HMM learning
- Las redes neuronales artificiales (Artificial neural networks): Architectures and sequence encoding, feed-forward algorithm, and back propagation

### I Semana: Introducción al curso

#### Teorías:

- Introducción al curso
- El sistema inmune
- UNIX crash course 101
- C-programming crash course 101

#### Trabajo Práctico:

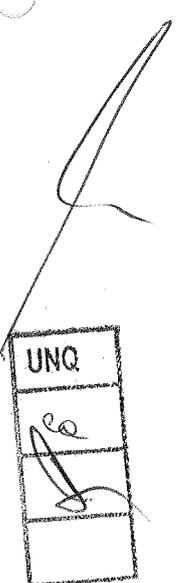
- C programming usando pointers y linked listas

#### Teorías:

- Estructura del curso, conceptos de Unix y de lenguaje c-programming.

#### Trabajo Prácticos:

- Comandos unix y desarrollo de un programa en C básico usando líneas de comandos.



## II Semana: Construcciones de "Weight matrix"

Teorías:

- Weight matrix (PSSM) construction,
- Gibbs sampling
- Psi-Blast

Trabajo Prácticos:

- PSSM usando secuencias pre alineadas. Corrección por pseudo cuentas y clustering secuencial.

Teorías:

- Información secuencial, motivos, pesado de secuencias, pseudocuentas y matrices.

Trabajo Prácticos:

- Implementar un programa para construcción de matrices y uso de pseudo cuentas.

## III Semana: Alineamiento secuencial.

Teorías:

- Matrices Blosum
- Alineamiento secuencial
- Blast y Psi-Blast

Trabajo Prácticos:

- Implementación de Smith-Waterman Dynamic Programming

Teorías:

- Base teórica para el estudio de matrices tipo BLOSUM y introducción detallada a algoritmo de dynamic programming y su aplicación a alineamiento secuencial. En particular se verá cómo y porqué el uso de perfiles secuenciales mejora sustancialmente la calidad de los alineamientos secuenciales.

Trabajo Prácticos:

- Implementación de un programa para aplicar Smith-Waterman dynamic programming en  $O^3$  y tiempo  $O^2$ .

#### IV Semana: Selección de datos y redundancia

- Teorías:
- Bases de datos biológicas.
- Redundancia de datos y algoritmos para su reducción (Hobohm1 and Hobohm2).

Trabajo Prácticos:

- Implementación de algoritmos para reducción de homología.

Teorías:

- Se explica porque los algoritmos de reducción de redundancia son esenciales para el desarrollo de distintos algoritmos de predicción bioinformática. Los algoritmos más usados son Hobohm 1 and Hobohm 2.

Trabajo Prácticos:

- Implement program for Hobohm 1 and Hobohm 2 data redundancy reduction

#### V Semana: Hidden Markov Models

Teorías:

- Introducción a los modelos Hidden Markov
- Decodificación Viterbi
- Algoritmo Forward/Backward, decodificación posterior, algoritmo Baum-Welsh
- Profile Hidden Markov Models.

Trabajo Prácticos:

- Implementación de Viterbi y decodificación a posteriori
- Uso del HMMER para entrenar un Hidden Markov model



Teorías:

- Se introduce el concepto de hidden Markov models. Se describe como se construyen los modelos a partir de los datos, y se muestra el uso de Viterbi y de decodificación a posteriori. También se muestra como HMM puede ser usado para predecir aspectos en proteínas y secuencias.

Trabajo Prácticos:

- Desarrollo de programas implementando Viterbi y decodificación posteriori usando HMM. Uso de HMMER para construir perfiles y su evaluación.

### VI Semana: Redes neuronales artificiales.

Teorías:

- Introducción
- Codificación secuencial
- Algoritmo Feed forward
- Back-propagation y entrenamiento de la red
- Network training - cross-validation
- Web Trabajo Practicos

Trabajo Prácticos:

- Implementación de sequence encoding feed forward y back-propagation algorithms

Teorías:

- Redes neuronales artificiales (ANN). Se describe como usar información biológica para codificar y entrenar ANN. Uso de algoritmos de feed-forward. Uso de redes neuronales para la predicción biológica. Estudio del algoritmo de back-propagation para alimentación y entrenamiento. Problema de sobrefito. Validación y entrenamiento de redes.

Trabajo Prácticos:

- Desarrollo de programas que implementen el algoritmo de forward and back-propagation para entrenamiento y evaluación de ANN.



## VII y VIII Semana: Trabajo final

En este periodo los estudiantes aplicaran uno o más algoritmos desarrollados en el curso para estudiar un determinado problema biológico. En el mismo se podrá desarrollar métodos para la predicción o simplemente la aplicación de programas para resolver alguna problemática. El proyecto deberá documentarse como un informe de la forma de los manuscritos científicos.

## **Examen final**

### Bibliografía general

Todos los papers y fotocopias será entregado a los alumnos al inicio del curso.

- Immunological Bioinformatics. Lund et al. MIT Press. Focus on chapters 3-4, but other chapters of the book are also relevant.
- Tom Macke. C programming notes.
- A General method applicable to search for similarities in the amino acid sequences of two proteins. S. B. Needleman and C. D Wunsch J.Mol. Biol (1970), 48.
- Identification of common molecular subsequences. T. F. Smith and M. S. Waterman. J. Mol. Biol. (1981), 147.
- An improved algorithm for matching biological sequences. O. Gotoh. J. Mol. Biol. (1982), 162
- Selection of representative protein data sets. U. Hobohm et al. Protein Science. 1992.
- Protein distance constraints predicted by neural networks and probability density functions. O. Lund et al. Protein Engineering. (1997).
- Hidden Markov Models notes by Anders Krogh



**Metodología:** Teórico-práctico

**Modalidad:** Presencial

**Requisitos de asistencia:** Asistencia al 80 % del total de las clases.

**Evaluación:** Examen oral e informe escrito. Ponderación, informe escrito (50%) y examen oral (50%).

**Certificación:** Certificados de Asistencia y Aprobación de la Universidad Nacional de Quilmes.

**Cupo máximo:** 20 alumnos.

**Arancel:**

Arancel general de \$ 960.-

Los egresados de la Universidad están exentos del pago.

**Presupuesto:**

La realización del curso quedará sujeta a que la recaudación de fondos garantice la cobertura de su presupuesto.



**Requerimientos:**

Los currículos de los docentes constan de fs. 9 a 36 del Expediente citado en el Visto.

ANEXO RESOLUCIÓN (R) N°:

-00108

Mg. Alfredo Alfonso  
Secretario General  
UNIVERSIDAD NACIONAL DE QUILMES

Gustavo Eduardo Lugones  
Rector  
Universidad Nacional de Quilmes