

PROTOCOLO

Redes Neuronales para la Minería de Datos y Textos: Aplicación al Análisis Exploratorio y Descubrimiento de Conocimiento en Grandes Bases de Datos de Información Biomédica.

Introducción y Motivación

- *Minería de Datos y Textos.*
- *Redes Neuronales.*
- *Redes y Mapas Auto-organizados.*
- *Visualización de Información.*
- *Bibliometría, Cienciometría e Informetría..*
- *Implicaciones, Conclusiones y Perspectivas.*
- *Bibliografía*

Participantes de México.

Participantes de Cuba.

Antecedentes de la Colaboración

Objetivo General

Metas

Productos

Aplicaciones

Tareas

Importancia del Intercambio de Investigadores para el Desarrollo de la Propuesta Infraestructura, Equipo y Apoyo Técnico Disponibles (LDNL-Mex.)

Información de la Contraparte Extranjera

- **Antecedentes de la Institución Extranjera**
- **Objetivos de la Institución Extranjera (en el área específica del proyecto).**
- **Experiencia en el Área a Desarrollar.**
- **Infraestructura, Equipo y Apoyo Técnico Disponibles.**

Programa de Trabajo (2003)

Anexo 1 Aplicaciones

Anexo 2 Premio de la Academia de Ciencias de Cuba

Anexo 3 Pruebas Preeliminares de Eficiencia (“Bench Mark Test”)

PROTOCOLO

Redes Neuronales para la Minería de Datos y Textos: Aplicación al Análisis Exploratorio y Descubrimiento de Conocimiento en Grandes Bases de Datos de Información Biomédica.

Introducción y Motivación

Minería de Datos y Textos. La tecnología moderna permite la creación de grandes almacenes de datos (crudos) que requieren ser explorados en búsqueda de información refinada (conocimiento). Un reto tecnológico de actualidad es desarrollar métodos y procedimientos “automáticos” para procesar estos datos y convertirlos en conocimiento útil para la toma de decisiones. A esta área de trabajo se le ha dado en llamar Minería de Datos (también “Exploratory Data Analysis” o “Knowledge Discovery”). Muchos problemas de la informática actual requieren de la aplicación de técnicas de esta disciplina. Basándose en diversos métodos matemáticos y en nuevos paradigmas de la inteligencia artificial (e.g.: redes neuronales artificiales), la minería de datos ha emergido como un campo estratégico de la informática moderna. Con estas nuevas tecnologías la minería de datos ha tenido un importante impacto en la industria y en el mundo de los negocios, para resolver una gran variedad de problemas como: planeación económica, inteligencia empresarial, finanzas, análisis de mercados y análisis de perfiles de clientes. Recientemente ha surgido una rama más especializada: la Minería de Textos, que está siendo útil para descubrir conocimiento y organizar creativamente información digital en diversos campos de conocimiento.

Redes Neuronales. Las redes neuronales sirven para realizar automática y eficientemente múltiples tareas como por ejemplo: modelación, optimización, regresión, clasificación, “clustering”, memorización, aprendizaje asociativo, control adaptativo, pronóstico y predicción de series de tiempo. Por esto constituyen un recurso estratégico de la tecnología informática actual. La minería de datos es uno de los campos de aplicaciones de la redes neuronales. En correspondencia con la diversidad de tareas, existen muchos tipos y paradigmas diferentes de redes neuronales; cada problema particular requiere un paradigma y el diseño de una arquitectura apropiada. Uno de los problemas clave de investigación es la selección de un modelo óptimo.

El aprendizaje o proceso de entrenamiento de una red neuronal se entiende como la etapa mediante la cual la red neuronal adquiere la capacidad de desempeñar las funciones específicas

que requiere el problema que se pretende abordar. Los distintos modelos se clasifican de acuerdo al paradigma básico de aprendizaje que estos usan. Los tres principales paradigmas de aprendizaje son: supervisado, no supervisado y de refuerzo. El aprendizaje supervisado es el paradigma de entrenamiento más común para realizar tareas de clasificación y aplicaciones de predicción. En este enfoque el algoritmo de aprendizaje ajusta adaptativamente la diferencia entre el output deseado y el que la red neuronal genera. Así la red aprende a realizar automáticamente una tarea prescrita. Esto requiere que el diseñador de la red conozca de antemano la respuesta de la red en una muestra de casos que son usados para el entrenamiento. En cambio el aprendizaje no supervisado es apropiado para resolver problemas en los que se tiene clara la pregunta pero no se conoce la respuesta. Por ejemplo la pregunta puede ser: ¿Cómo se relacionan los datos? ¿Cuáles datos son similares y en que consiste esta similitud? ¿Qué patrones o estructuras globales subyacen en la colección de datos? Así pues, el cálculo que las redes neuronales no supervisadas realizan automáticamente termina revelando una organización o estructura final. Por esta razón, a este tipo de redes neuronales son llamadas redes “auto-organizadas”. Estas redes son muy socorridas para realizar tareas de clustering y segmentación jerárquica de datos, que son funciones de gran utilidad para el descubrimiento de conocimiento en grandes bases de datos.

Redes y Mapas Auto-organizados. Las redes que ocupan un paradigma de entrenamiento no supervisado, se basan en un enfoque competitivo: durante el proceso de entrenamiento cada uno de las neuronas o nodos de la red compiten entre sí para determinar cual de ellos tiene mayor capacidad de representar cada uno de los datos que se le van presentando a la red. Esta competencia va generando progresivamente una reorganización (interna) de las neuronas que finalmente puede ser representada en diversos tipos de mapas bidimensionales que permiten revelar visualmente la organización global (del conjunto de datos de entrada) encontrada por la red. Dentro de las redes neuronales de este tipo más usadas se encuentran: “Adaptive Resonance Theory” (ART) y los “Features Maps” de Kohonen (SOM).

Visualización de Información. Uno de los objetivos de la informática moderna es poner la alta tecnología de graficación al servicio de la percepción de información para facilitar el proceso de interiorización del conocimiento. Una medida de la novedosa emergencia de esta problemática nos lo da la reciente aparición (en este 2002) de la revista: *Information Visualization*.

En el terreno de la tecnología de visualización, las redes neuronales son muy apreciadas pues proveen un medio adecuado para la producir mapas topográficos y otras formas de representación o “metáforas visuales”.

Bibliometría, Cienciometría e Informetría. La Bibliometría aprovecha la cuantificación de la información para diferentes fines. Su actual denominación tiene escasamente treinta años. En la actualidad, la Bibliometría se ha beneficiado de técnicas conceptuadas en la “minería de datos”, donde, a las herramientas estadísticas ya establecidas, se suman los más modernos enfoques de análisis multivariado tales como las “redes neuronales artificiales”. Por una parte, tiene aplicaciones relacionadas con la gestión y el uso de la información en bibliotecas, donde surgió. Por otra parte, de la Bibliometría se han derivado aplicaciones en el campo de la política científica, que es lo que se conoce como Bibliometría Evaluativa. Más recientemente la Bibliometría también se ha ocupado de la vigilancia tecnológica, como apoyo a la toma de decisiones en ambientes empresariales (producción y servicios). También ha influyó al nacimiento de otra especialidad, vinculada al estudio de la producción científica y la organización de la ciencia, la llamada Cienciometría. Ambas, han dado lugar a otra disciplina todavía más reciente, la Informetría.

Implicaciones, Conclusiones y Perspectivas. A pesar de todos los avances recientes el análisis automático de información digital, y la tecnología para producir representaciones visuales, se encuentra todavía en un estadio preliminar. Esto se manifiesta en el surgimiento de una serie de metodologías y técnicas, que aunque promisorias, no tienen un gran sustento teórico y su validación es todavía demasiado empírica o limitada al ámbito de bases de datos y/o problemas que son estudiados con indicadores específicos.

Por otra parte, este escenario de novedosa aplicación tecnológica de conocimientos científicos, ha ofrecido también una oportunidad de negocio que está siendo aprovechada por diversas compañías que se dedican a producir diversas herramientas de software que sirven para explorar y visualizar creativamente el conocimiento encontrado en la constelación de bancos de información digital que hoy tenemos disponibles. Como hemos apuntado anteriormente estas herramientas implican sofisticados métodos de procesamiento matemático y el estado del arte en técnicas de visualización e ingeniería de software. En consecuencia estas herramientas o servicios suelen ser muy costosos (demasiado para las instituciones típicas de nuestros países). Por ejemplo, con los presupuestos disponibles, nuestro grupo ha podido solamente tener acceso a uno de estos sistemas (*Viscovery SOMine*), gracias a la donación otorgada por la compañía productora

(*Eudaptics, GmbH; Austria*), en virtud de las investigaciones y aplicaciones que, con una versión de prueba de este sistema habíamos realizado.

Bibliografía

- Saavedra, Oscar; Sotolongo, Gilberto, Guzmán, MV. Mapeo autoorganizado de las revistas científicas y técnicas de América Latina y el Caribe. Aceptado. ACIMED.
- Guzmán, MV., Sotolongo, G. Mapas Tecnológicos para la Estrategia Empresarial. El caso de la Neisseria meningitidis. Aceptado. ACIMED.
- Saavedra, O.; Sotolongo, G.; Guzmán, MV. Medición de la producción científica en América Latina en el campo agrícola y afines: un estudio bibliométrico. Revista Española de Documentación Científica. 2002; 25(2):151-161.
- Sotolongo, G; Guzmán, MV. Saavedra, O.; Carrillo, H. Mining Informetrics Data with Self-organizing Maps. In: Davis, M.; Wilson, CS. (Eds.). Proceedings of the 8th International Society for Scientometrics and Informetrics. ISBN:0-7334-18201. Sydney, Australia July 16-20. Sydney: BIRG; 2001: 665-673.
- Sotolongo, G.; Guzmán, MV. Aplicaciones de las redes neuronales. El caso de la bibliometría. Ciencias de la Información. 2001; 32(1):27-34.
<http://www.idict.cu/cinfo99/v32n1a2001/bibliometria.htm>.
- Sotolongo, G; Suárez, CA.; Guzmán, MV. Modular Bibliometrics Information System with Proprietary Software (MOBIS-ProSoft): a versatile approach to bibliometric research tools. : Library and Information Science Electronic Journal (LIBRES). 2000; September 30, Volume 10 Issue 2. <http://libres.curtin.edu.au/>
- Macías-Chapula, C A; Sotolongo-Aguilar, G R; Magde, B; Solorio-Lagunas, J. Subject Content Analysis of AIDS Literature Produced in Latin America and the Caribbean. Scientometrics, 1999 ;46 (3) :563 – 574.
- Sotolongo, G; Suárez, CA.; Guzmán, MV. Modular Bibliometrics Information System with Proprietary Software: In: Macías-Chapula, C. (Ed.). Proceedings of the Seventh International Society for for Scientometrics and Informetrics. ISBN: 968-619091-0. Universidad de Colima, México. 1999; 313-322.

Bibliografía. (Otros autores)

- Berson, A and S.J. Smith. Data Warehousing and Data Mining. McGraw-Hill. 1997.

- Cabena, P.; Hadjnia, P.; et. Al. *Discovering Data Mining: From Concept to Implementation*. Prentice Hall, 1999.
- Bigus, J. *Data Mining With Neural Networks*. McGraw-Hill. 1996.
- Kohonen, T. *Self-Organizing Maps*. Springer. 2001.
- Kaski, S. *Data Exploration Using Self-Organizing Maps*. Thesis for the degree of Doctor of Technology. Helsinki University of Technology. 1997.
- Kohonen, Kaski, Lagus, Salojärvi, Honkela, Paatero and Saarela. Self Organization of a massive document collection. *IEEE TRANSACTIONS ON NEURAL NETWORKS*, vol. 11, No.3, May 2000
- Chaomei, Chen; et al. Visualizing and Tracking the Growth of Competing Paradigms: Two Case Studies. *JOURNAL OF THE AMERICAN SOCIETY FOR INFORMATION SCIENCE AND TECHNOLOGY*, 53(8): 678–689, 2002.
- Chaomei, Chen; Ray J. Paul. Visualizing a Knowledge Domain's Intellectual Structure. *Research Feature*. March 2001: 65-71.
- Deboeck, GS. Pattern Recognition and Prediction with Self-Organizing Map. Supporting Software Review: visualization thorough Viscosity. Gordian Institute Electronic Newsletter. [Http: //www.gordianknot.com](http://www.gordianknot.com). Acceso 20 de enero de 1999.
- G.S. Davidson et al., "Knowledge Mining with VxIn-sight: Discovery through Interaction," *J. Intelligent Information Systems*, vol. 11, no. 3, 1998, pp. 259-285.
- H. Small, "Visualizing Science by Citation Mapping," *J. Am. Soc. Information Science*, vol. 50, no. 9: 799-813. 1999.
- Kornikov, A.R.. Intelligent technologies new opportunities for modern industry. *Information Technology*, 3:1-14, 1997
- Clarie François, Xavier Polanco, Jean-Charles Lamirel *Information Visualization and Analysis for Knowledge Discovery: using a Multi Self-Organizing Mapping* 4th European Conference of Principles and practice of Knowledge Discovery in Databases (PKDD), Lyon, France, September 12-16 2000, Workshop: "Machine Learning and Textual Information Access", edited by Hugo ZARAGOZA, Patrick GALLINARI and Martin RAJMAN.
- Xavier Polanco, Claire François, Jean-Charles Lamirel *Using Artificial Neural Networks for Mapping of Science* Sixth International Conference on Science and Technology Indicators, The Netherlands, 24-27 May 2000.
- Polanco, Xavier; Francois, C., Keim, JF. Artificial neural network technology for the classification and cartography of scientific and technical information. In: Periz, B., Egghe, L.

Proceeding Sixth International Conference of the International Society for Scientometrics and Informetrics; Jerusalem, Israel, 1997. Jerusalem, Universidad Hebrew of Jerusalem, 1997, pp. 319-330.

- Schmoch, U. Evaluation of Technological Strategies of Companies by Means of MDS Maps. *IJTM*. 1995, 10(4,5,6):426-440.
- Swanson, DR.; Smalheiser, NR. An interactive system for finding complementary literatures: a stimulus to scientific discovery. *Artificial Intelligence*. 1997; 91:183-203.

Participantes de México.

Laboratorio de Dinámica no Lineal, Facultad de Ciencias,
Universidad Nacional Autónoma de México.

Responsable del proyecto. Humberto Carrillo Calvet

Técnico. Antonio Carrillo Ledesma

Tesista. (Licenciatura) Elio A. Villaseñor García

Pilar Ladrón Guevara

Participantes de Cuba.

Instituto Finlay, La Habana.

Responsable del proyecto. Gilberto R. Sotolongo Aguilar

María V Guzmán Sánchez

Romel Calero Ramos

Armando Acosta Domínguez

Antecedentes de la Colaboración

Desde hace cinco años la Dirección de Información Científica del Instituto Finlay, la Habana Cuba; y el Laboratorio de Dinámica no Lineal de la Facultad de Ciencias de la UNAM, México D. F. han estado intercambiando información e ideas sobre la problemática apuntada anteriormente. Esta colaboración ha implicado visitas mutuas, el diseño de experimentos y la realización de aplicaciones utilizando redes neuronales auto-organizados para la obtención de mapas de conocimiento. Este esfuerzo se ha sostenido y ha rendido primeros resultados:

1. El Dr. Carrillo participa como tutor de la tesis doctoral de G. Sotolongo;
2. Algunos de los resultados alcanzados conjuntamente han sido publicados en el artículo: "Mining Informetric Data with Self Organizing Maps", en las memorias del Congreso anual de la International Society for Scientometrics and Informetrics (Sydney, Australia 2001);

3. G. Sotolongo y M.V. Guzmán obtuvieron el Premio Anual 2001 de la Academia de Ciencia de Cuba por su trabajo: *"Herramienta para el Análisis de Oportunidades Científicas y Tecnológicas"*, teniendo al Dr. Carrillo como uno de los reconocidos colaboradores.

Para poder acelerar procesos y acortar tiempos en la realización del proyecto que ahora nos hemos planteado, requerimos una llevar a cabo una colaboración más estrecha e intensa.

Objetivo General

Ante la presente coyuntura tecnológica en el terreno de la Informática y los múltiples "modos de hacer", ante la existencia de decenas de sistemas software (en experimentación, parcialmente compatibles, con limitaciones funcionales, sin validación y costosos), como resultado de una evaluación de conveniencia y factibilidad, hemos constituido un grupo interdisciplinario que tiene por objeto apropiarse de estas nuevas tecnologías y ponerlas al servicio de nuestras comunidades científicas.

Así pues, nos proponemos agenciarnos, diseñar y construir herramientas teóricas y computacionales, basadas en métodos multivariados y adaptativos (e.g. redes neuronales), entrenarnos en su uso y aplicarlas para hacer análisis exploratorio de datos y textos para el descubrimiento de conocimiento en bancos digitales de información biomédica (en la perspectiva de la incipiente ciencia informétrica).

Dada la adscripción y la orientación profesional de los miembros del grupo, de manera natural, nuestro ámbito de aplicación y experimentación será el de la ciencia y tecnología biomédica.

Consideramos que este trabajo puede tener un sensible impacto local, en lo que se refiere a la planeación y la facilitación de la investigación biomédica y la producción de desarrollos biotecnológicos.

Metas

Esta etapa de nuestro proyecto tiene dos metas principales:

1. Desarrollar una metodología e integrar una colección de diferentes herramientas de software para constituir un sistema (modular) para la exploración el análisis y la búsqueda de conocimiento en las bases de información biomédica (artículos de investigación científica y patentes) que tenemos disponibles en la actualidad, e.g.: *Medline*.
2. En colaboración con miembros de la comunidad biotecnológica (cubana y/o mexicana), aplicar el sistema y la metodología desarrollada (en plan piloto) para el análisis de algunos casos de estudio de interés para nuestros países.

Productos.

Además de la metodología, el sistema modular y las aplicaciones, se producirán artículos de investigación y dos tesis de grado.

En el transcurso del primer semestre de 2003 debe quedar concluida la defensa (en Cuba) de la tesis doctoral de G. Sotolongo, bajo la tutoría de H. Carrillo, razón por la cual también se producirán intercambios y precisiones en tal sentido.

En el transcurso del segundo semestre de 2003 debe quedar concluida la defensa (en México) de la tesis de licenciatura de E. Villaseñor, bajo la tutoría de H. Carrillo.

Se presentarán también trabajos en Congresos Nacionales e Internacionales.

Aplicaciones.

En cuanto a las aplicaciones nos hemos planteado, durante el primer año, iniciar la exploración de la información relativa a dos temas en los que el LDNL-México y el IF-Cuba tienen un avanzado grado de especialidad:

A1. Vacunas contra la Tuberculosis. (Especialista en el tema: Dr. Armando Acosta Domínguez)

A2. Aplicación de modelos y métodos de la Dinámica no Lineal en las Ciencias Biomédicas.
(Especialista en el tema: Dr. Humberto Carrillo Calvet)

Como hemos apuntado, las redes neuronales aplicadas al análisis y visualización de Información es un campo emergente en un mundo cada vez más agobiado por el “exceso” de información. El

propósito de este proyecto es que el sistema diseñado pueda ser aplicado como una herramienta de análisis con múltiples propósitos. Algunos ejemplos de mapas de conocimiento se presentan en el Anexo 1.

En un estudio sobre la situación de un mercado más general (considerando las funciones que estarían incluidas en el sistema propuesto), se estima un potencial de aplicación vinculado a: finanzas en general, economía en general, estudios de mercado y clientes (Marketing), bioinformática, gestión del conocimiento, prospectiva tecnológica, investigación científica e innovación, inteligencia competitiva, gobierno y ministerios.

Como botón de muestra de la utilidad que estas técnicas pueden tener, en el Anexo 3, se presenta un reporte reciente de la Oficina Cubana de la Propiedad Industrial, evaluando (positivamente) algunos resultados preeliminares obtenidos aplicando la metodología de nuestro grupo.

Tareas.

La realización de las metas planteadas implica el cumplimiento de una mínima secuencia de tareas:

1. Identificar y conocer el objeto de estudio. Familiarizarse con las diferentes estrategias de búsqueda, posibles fuentes de información y características de diferentes motores de búsqueda, etc.
2. Búsquedas y salvar descarga de ficheros resultantes.
3. Tratamiento (conversión) del fichero resultante. Importar ficheros hacia una base de datos.
4. Creación de la base de datos.
5. Normalización de la base de datos. Depuración o filtraje de ruidos y datos erróneos.
6. Manipulación de la información contenidas en la base de datos para el tratamiento.
7. Preprocesamiento matemático (transformaciones, obtención de indicadores), escalado multidimensional (de los diferentes campos) y selección de la forma de representación numérica adecuada para la manipulación estadística de los datos y/o el entrenamiento de la red neuronal. Selección de las funciones de distancia y costo apropiadas en el espacio multidimensional para la preservación de la organización topológica de los datos, en los mapas topográficos.
8. Visualización de los resultados (clusters) en forma de mapas.
9. Evaluación de la utilidad de diferentes algoritmos para hacer análisis de “clustering” (para la clasificación y organización de los datos) con diferentes tipos de problema.

10. Evaluación de la utilidad de varias técnicas de clasificación, determinación de sombreados y contornos de los clusters. (para la clasificación y organización de los datos) con diferentes tipos de problema.
11. Interpretación y análisis comparativo de la información contenida en los distintos mapas. (Este es el momento del descubrimiento de conocimiento).
12. Crítica y Validación por especialistas.
13. Elaboración de conclusiones y recomendaciones.

El logro fundamental de este proyecto, es lograr una plataforma automatizada que garantice una secuencia lógica de todos los resultados, desde el primer dato salvado hasta la posterior representación visual en un mapa. Todo lo anterior permitiría enriquecer el procesamiento, visualización y análisis de los indicadores bibliométricos, cientiométricos y patentométricos con una metodología propia (UNAM-Finlay) y a bajo costo (los sistemas comerciales para estos propósitos, tienen un costo de entre 10 mil a medio millón de dólares). Esto implica que en la región, tanto universidades, organismos públicos de investigación y diferentes empresas de la región puedan tomar decisiones más acertadas y evaluar el desempeño de proyectos, instituciones, regiones o países. Por otra parte los Ministerios y Gobiernos no escapan del constante crecimiento de la información y de la influencia de los medios de divulgación en la opinión pública nacional e internacional. Estas entidades necesitan cada día más monitorear y vigilar diferentes datos relevantes del país o de un sector científico-tecnológico. A partir del sistema propuesto se podrán hacer inferencias sobre la dependencia tecnológica de un país, región, institución, sector económico hacia otras tecnologías, países o instituciones. Se logra, además, determinar los términos de referencia de los umbrales de los indicadores sobre el desempeño científico de los investigadores del sector de interés y su comparabilidad internacional. Se le adiciona a lo anterior, la posibilidad de evaluar el estado de la tecnología en un área o sector determinado, vislumbrar problemas económicos o financieros, conocer el estado de opinión de un hecho significativo o de una persona, y todo aquello relacionado con el análisis de los recursos intangibles.

Especialistas de otras instituciones como por Ejemplo la Oficina Cubana de la Propiedad Industrial de Cuba (OCPI) estiman que para el análisis de las referencias bibliográficas de unas 1200 patentes invertirían en el procesamiento manual de la información unos tres meses, mientras que con un sistema automatizado sólo invierten aproximadamente una semana. La optimización del tiempo de respuesta permite a este tipo de empresa y otras consultoras establecer servicios de valor añadido, donde brindan “información inteligente sobre productos emergentes, líneas de

investigación, posibles licenciamientos, etc. Así mismo se ha pensado que este sistema podrá ser implementado en Observatorios de Ciencia y Tecnología (a nivel macro y micro) y dentro de los sistemas de vigilancia científico tecnológica de las empresas e instituciones.

Importancia del Intercambio de Investigadores para el Desarrollo de la Propuesta

La parte cubana de nuestro grupo tiene experiencia en el campo de la investigación bibliométrica. La parte mexicana tienen experiencia en diversas aplicaciones de modelos no lineales y métodos matemáticos adaptivos (e.g.: redes neuronales) y al desarrollo de software para aplicaciones científicas.

Las estancias en el extranjero (y las visitas recíprocas de los participantes cubanos) están programadas para mutuamente transferirnos conocimiento y experiencia, así como también planear y discutir los avances el trabajo y afinar la escritura de los reportes de las investigaciones en curso.

Se han planeado estancias de veintiún días para aprovechar al máximo cada viaje.

Infraestructura, Equipo y Apoyo Técnico Disponibles (LDNL-Mex.)

Apoyo Técnico

- Técnico Tiempo Completo P. Mat. Antonio Carrillo Ledesma

Equipo de Cómputo

- 5 PC, 1 Notebook y 1 Estación de Trabajo Silicon Graphics
- Red Ethernet 10/100 en cada equipo
- 2 Impresoras Láser
- 1 Impresora de inyección de tinta

Las características del equipo se detallan a continuación

- 1 PC 166 Mhz con Windows 98
- 1 PC 100 Mhz con Linux
- 1 PC 450 Mhz con Windows 2000 Impresora HP 1100
- 1 PC 500 Mhz con Windows 2000 Impresora DJ 875

- 1 PC 933 Mhz con Windows 2000 Impresora HP 1000
- 1 NoteBook 300 Mhz con Windows 98
- 1 WS Silicon Graphics

INFORMACION DE LA CONTRAPARTE EXTRANJERA.

Antecedentes de la Institución Extranjera. El Instituto Finlay es una organización científica que se dedica a la investigación y producción de vacunas. Es una institución al servicio de la salud, heredera y continuadora de más de cien años de tradición científica.

Está considerada en su rama como una de las de mayor prestigio y reconocimiento a nivel internacional.

Ser creadores y productores de la primera y única vacuna efectiva contra el meningococo del grupo B (VA-MENGOC-BC <cartera/vamengoc.htm>®) fue sólo el inicio de una larga meta en nuestra estrategia por encontrar nuevos fármacos contra enfermedades que aún carecen de una medida profiláctico-curativa.

Trabajar en la producción de vacunas clásicas, especialmente en las del Programa Ampliado de Inmunizaciones (PAI), y lograr el desarrollo de vacunas combinadas es otra de las líneas de trabajo que sitúan al Instituto Finlay en la vanguardia mundial de la industria biotecnológica y médico-farmacéutica.

Los más de 900 trabajadores que conforman el mayor capital del Instituto Finlay poseen reconocida experiencia en el campo de las vacunas. Más del 70% de ellos poseen alta calificación técnico-profesional, mientras que su promedio de edad no rebasa los 36-37 años. Las mujeres (cerca del 50%) ocupan un lugar destacado en este importante universo científico.

Este colectivo de especialistas realiza sus actividades fundamentales en las áreas de investigación, desarrollo, producción, control de la calidad, servicios ingenieros, asistencia técnica aplicada, servicios bibliotecarios e informativos, marketing y docencia de pre y pos grado.

Muchos de ellos poseen más de 20 años de experiencia en estas especialidades lo cual garantiza la preparación y continuidad de los proyectos con la incorporación de nuevos talentos egresados de las universidades.

Objetivos de la Institución Extranjera (en el área específica del proyecto). La Dirección de Información Científica del Instituto Finlay, subordinada a la Vicepresidencia de I+D de Productos tiene como objetivos en el área específica del proyectos, el desarrollo de métodos y herramientas

para el análisis de información científica contenida en bases de datos; así como la realización de análisis en tópicos específicos de interés institucional.

Experiencia en el Area a Desarrollar. Con sus antecedentes desde 1990, la actual Dirección de Información Científica se conformó alrededor del CIBI Finlay (Centro de Información y Biblioteca del Instituto Finlay). La Dirección como tal se creó en 1995 y prácticamente desde 1993 ha venido trabajando en los objetivos enunciados anteriormente. Un ejemplo elocuente de la experiencia en el área a desarrollar lo constituye el hecho de que en 2001, su director Gilberto Sotolongo y su más cercana colaboradora María V Guzmán, fueron acreedores del Premio Anual 2001 de la Academia de Ciencia de Cuba (“Herramienta para el Análisis de Oportunidades Científicas y Tecnológicas”), conjuntamente con un grupo de colaboradores externos entre los que se encuentra el Dr. Humberto Carrillo (Director del Laboratorio de Dinámica no Lineal de la Facultad de Ciencias de la UNAM).

Infraestructura, Equipo y Apoyo Técnico Disponibles. En Cuba el proyecto se desarrollaría bajo la dirección de Gilberto Sotolongo con la participación de parte del equipo de trabajo del Grupo de Gestión de Información dirigido por María V. Guzmán y también con la participación de Romel Calero. Adicionalmente se contará con el apoyo del Dr. Armando Acosta, Investigador Titular y Director del Departamento de Biología Molecular dependiente de la Vicepresidencia de Investigaciones del Instituto Finlay.

Para el desarrollo del trabajo se cuenta con dos PC IBM compatibles del Modelo Pentium 2 a 133Mh (4 Gb y 10 Gb de disco duro y respectivamente y 64 Mb de RAM. Equipamiento que se considera insuficiente. También se cuenta con diferentes software útiles para la realización de los análisis de información previstos.

Programa de Trabajo (2003)

En el primer semestre se prevé una estancia de un especialista del Laboratorio de Dinámica no Lineal (LDNL), visitando la Dirección de Información Científica (DIC) Cuba.

En el segundo semestre, un especialista de la Dirección de Información Científica (DIC) visitando el LDNL. Ambas estancia será de tres semanas.

Durante el primer semestre se deben producir intercambios de conocimientos (vía E-mail y otras formas de comunicación) para instalar y probar las herramientas de trabajo, poner a punto la metodología de trabajo, así como discutir el diseño de la nueva metodología y el enfoque con el que se van a abordar las aplicaciones sobre las que se va a trabajar conjuntamente. Este período contempla la visita de un especialista del LDNL, visitando la DIC, para realizar los ajustes pertinentes y enriquecer el intercambio de información. En esta etapa se formularán los detalles del plan de trabajo para el segundo semestre.

También en el transcurso del primer semestre debe quedar concluida la defensa (Universidad de la Habana) de la tesis doctoral de G. Sotolongo, bajo la tutoría de H Carrillo, razón por la cual también se producirán intercambios y precisiones en tal sentido.

En el transcurso del segundo semestre se llevarán a cabo las aplicaciones programadas e iniciadas en el primer semestre. En esta etapa debe quedar redactada los reportes de las aplicaciones y una memoria de la metodología desarrollada que sirva de base para facilitar desarrollos informáticos ulteriores. También debe quedar concluida la tesis para obtener el grado de Matemático (UNAM) de E. Villaseñor, bajo la tutoría de H. Carrillo.

Al concluir este período, deben quedar también definidas las etapas subsiguientes para dar continuidad al proyecto durante el 2004.