

Herramientas computacionales para análisis epidemiológico multifactorial

Satu Elisa Schaeffer, 096633

*Facultad de Ingeniería Mecánica y Eléctrica
Universidad Autónoma de Nuevo León*

PAICYT 2015

Resumen

El proyecto propuesto plantea el desarrollo de herramientas computacionales para facilitar el estudio epidemiológico — modelado y análisis — multifactorial, donde datos sobre las cantidades de casos reportados de distintas enfermedades en México se combinan con información sobre las condiciones climáticas para identificar diferentes tipos de efectos del cambio climático en las enfermedades bajo vigilancia epidemiológica de la Secretaría de la Salud y además explorar efectos geográficos (si tasas de enfermedades en estados vecinos tienen una relación estadísticamente significativa) y proponer modelos que incorporen estos factores y permitan la introducción de factores adicionales.

Se cuenta con fuentes de información cuyo preprocesamiento (digitalización y estructuración) y combinación dentro del proyecto propuesto permitirá una visualización de información no accesible actualmente para investigadores mexicanos del área de la salud, abriendo puertas para la formulación de nuevos modelos para analizar y pronosticar epidemias, lo que permite en un largo plazo nuevas estrategias para la salud pública. Existen ya varias herramientas en prototipos preliminares que han sido desarrollados a lo largo del presente año en preparación para el proyecto propuesto, lo que facilitará la conclusión exitosa y puntual de este trabajo. Se contempla continuar el trabajo desarrollado en este proyecto en un futuro cercano con financiamiento externo.

1. Introducción

El cómputo científico es un área multidisciplinaria de las ciencias computacionales que combina elementos de matemáticas puras y aplicadas, que permite la automatización de muchas tareas de análisis de resultados experimentales de distintos campos de ciencias exactas, naturales y aplicadas. En particular en el área de la salud el impacto del procesamiento de datos por computadora ha permitido avances rápidos e impactantes.

El proyecto propuesto es una colaboración entre la *Facultad de Ingeniería Mecánica y Eléctrica* (FIME) y el *Centro de Investigación y Desarrollo en Ciencias de la Salud* (CIDICS) de la UANL para explorar áreas de investigación de la salud que pudieran beneficiarse del uso de técnicas de cómputo científico, iniciativa del Dr. Carlos E. Medina de la Garza, el director del CIDICS. Este proyecto es en sí un piloto tipo abre-puertas para futuras colaboraciones y se enfoca en el área de epidemiología. Su inspiración proviene de estudios previos de enfermedades respiratorias [2] del grupo de investigación del Dr. Gerardo Velasco del CIDICS donde se ha capturado manualmente información epidemiológica y climática para identificar posibles relaciones. El proyecto propuesto busca facilitar este tipo de investigación por automatizar fases de extracción y procesamiento de datos sobre casos diagnosticados de enfermedades y facilitar su combinación con factores adicionales tales como datos climáticos, geográficos y geopolíticos, igual como proveer herramientas de modelado y análisis de los mismos.

La *Organización Mundial de la Salud* vigila de forma sistemática enfermedades contagiosas como aquellas causadas por los virus de la influenza¹ desde hace más de medio siglo. A través del monitoreo puede preparar recomendaciones y estrategias de alerta para buscar controlar epidemias y así evitar pandemias. También en México se ha puesto atención en el impacto de la influenza [14]. De hecho, México cuenta con planes dedicados a monitoreo y control de epidemias de influenza [5] donde se enfatiza la importancia de “los lazos entre los tomadores de decisiones y el sector académico nacional” igual como la importancia de colaboración para lograr respuestas eficaces conjuntas en el caso de una emergencia epidemiológica. Son esfuerzos como estos los que recibirán los beneficios del proyecto propuesto.

2. Antecedentes

Las enfermedades cuyo contagio se considera peligroso para la población, como por ejemplo la influenza aviar [1] y aquellos transmitidos por mosquitos [8], han sido objeto de estudio intenso en el área de las ciencias de la salud por décadas, el modelado matemático [4] y el estudio estadístico de las tasas de ocurrencia [12] han sido las herramientas básicas. La propagación de una epidemia claramente no depende únicamente del número de pacientes sino también puede ser afectado por factores que no son médicos tales como densidad poblacional, número de personas expuestas, nivel de bienestar y desarrollo social, clima y ubicación.

Existen trabajos de análisis epidemiológico tomando en cuenta la ubicación geográfica del punto de infección [7, 10] y también se ha establecido que condiciones climáticas tienen un efecto en diferentes tipos de enfermedades infecciosas [2, 3]. Ya existen en la literatura modelos epidemiológicos desde hace décadas [6, 11], las aplicaciones más esenciales son el pronóstico de epidemias y el diseño de mecanismos de controlarlas (tal como campañas eficientes de vacunación y eliminación de hábitat de animales que propagan la enfermedad). Uno de los retos de lograr un impacto en la salud es obtener y combinar los datos de entrada que se requiere para aplicar estos modelos con facilidad, precisión y rapidez. Otro punto de interés es la aplicación de los avances en la ciencia de redes (inglés: network science) que ha avanzado notablemente en los últimos 20 años.

La *Secretaría de Salud* (a nivel federal), a través de la *Dirección General de Epidemiología* (DGE) publica semanalmente un *boletín epidemiológico* en formato PDF²; su repositorio de boletines históricos cuenta con boletines disponibles para la descarga desde el año 1981. Desafortunadamente no proveen los datos como hojas de cálculo para poder analizarlos. Los boletines semanales incluyen numerosas enfermedades, identificadas por su número de clasificación internacional [9] y se proporcionan las cantidades de casos diagnosticados por entidad federal (i.e., estado de la república); se muestra un ejemplo en la figura 1 en la página 3.

Investigadores del CIDICS han estado capturando esta información para algunas enfermedades respiratorias manualmente en hojas de cálculo, lo que implica mucho trabajo y además es un proceso propenso a errores al copiar datos numéricos visualmente de un archivo a otro (por su formato, copiar y pegar información no es una opción viable ya que se desacomodan las cifras al intentar copiar más de una celda del cuadro en el boletín a la vez). Con una herramienta para digitalizar todos estos datos epidemiológicos para todas las enfermedades y combinarlos con datos públicamente disponibles sobre clima, geografía y poblaciones abriría numerosas direcciones para investigación epidemiológica y facilitaría notablemente el trabajo de investigadores del área. Este proyecto propone la creación y difusión de herramientas para extraer, procesar, combinar y analizar información relevante para modelos epidemiológicos en México.

¹http://www.who.int/influenza/surveillance_monitoring/updates/latest_update_GIP_surveillance/en/

²http://www.epidemiologia.salud.gob.mx/dgae/boletin/intd_boletin.html

CUADRO 7.2 Casos por entidad federativa de Enfermedades Transmitidas por Vector hasta la semana epidemiológica 20 del 2014

ENTIDAD FEDERATIVA	Fiebre por Dengue CIE-10 ^a REV. A90				Fiebre Hemorrágica por Dengue CIE-10 ^a REV. A91				
	2014			2013	En Estudio	2014			2013
	Confirmados		Acum.	Confirmados		Confirmados		Confirmados	
	Sem.	Acum.		Acum.	Sem.	Acum.		Acum.	
	M	F				M	F		
Aguascalientes	-	-	-	-	-	-	-	-	-
Baja California	-	-	1	-	-	-	-	-	-
Baja California Sur	6	67	58	46	6	2	8	1	1
Campeche	6	50	56	51	5	-	10	8	19
Coahuila	-	-	1	3	3	-	-	-	-
Colima	2	49	69	269	26	-	28	21	92
Chiapas	12	65	73	184	21	19	95	73	240
Chihuahua	-	-	-	-	-	-	-	-	-
Distrito Federal	-	-	-	-	-	-	-	-	-
Durango	1	1	-	1	-	-	-	-	-
Guanajuato	-	-	-	-	-	-	-	-	-
Guerrero	17	130	105	473	71	9	74	74	296
Hidalgo	1	5	2	13	-	-	-	-	1
Jalisco	5	85	123	156	50	2	45	40	17
México	-	-	-	9	3	-	-	-	-
Michoacán	1	46	52	397	1	-	2	1	35
Morales	4	23	31	12	52	4	11	13	19
Navarro	6	21	37	53	173	3	48	50	121
Nuevo León	46	157	329	413	1	-	1	2	2
Oaxaca	5	49	68	176	40	7	50	46	83
Puebla	8	29	26	157	-	-	-	-	1
Querétaro	-	1	-	-	-	-	-	-	-
Quintana Roo	3	51	49	284	62	-	39	37	213
San Luis Potosí	-	24	28	157	-	-	1	1	7
Sinaloa	2	115	135	68	477	6	151	143	19
Sonora	7	43	40	23	17	-	16	13	1
Tabasco	4	27	51	792	7	-	9	16	200
Tamaulipas	4	19	19	101	8	-	2	4	37
Tlaxcala	-	-	-	-	-	-	-	-	-
Veracruz	2	63	92	328	56	1	53	42	276
Yucatán	4	103	111	188	40	2	61	68	102
Zacatecas	-	-	-	-	-	-	-	-	-
TOTAL	146	1 223	1 556	4 354	1 119	55	704	653	1 782

FUENTE: SINAVE/DGE/Saúd 2014. Información preliminar.

Figura 1: Un ejemplo de un cuadro de informe de casos identificados en un boletín.

3. Objetivos y metas

La *hipótesis* de este trabajo es la siguiente: la digitalización de datos epidemiológicos mexicanos a gran escala y su combinación con bases de datos existentes habilita una exploración sistemática, profunda y estadísticamente robusta de efectos que distintos factores (climáticos y socioeconómicos, por ejemplo) tienen en distintos tipos de enfermedades.

El *objetivo principal* es creación de herramientas computacionales para la extracción y la combinación exitosa de datos epidemiológicos, climáticos, geográficos y geopolíticos igual como su modelado y análisis iniciar con técnicas existentes. Cumpliendo con esto, se podrá proceder a solicitar un proyecto de marco mayor, como por ejemplo un proyecto de ciencias básicas de CONACyT, para derivar modelos, pronósticos y estrategias preventivas a partir de los datos obtenidos a través de este proyecto y el análisis preliminar de los mismos.

Objetivos específicos

- Digitalización de cuadros de tasas de incidencia semanales por entidad federativa de los boletines.
- Sincronización de los datos semanales epidemiológicos con los datos diarios climáticos.
- Construcción de series de tiempo multifactoriales semanales por entidad federativa combinando los datos epidemiológicos y los datos climáticos.
- Construcción de un modelo estadístico — espacial y temporal — aplicando la teoría de grafos en la relación de vecindad entre las entidades federativas.
- Clasificación de enfermedades según las características de sus modelos con técnicas de agrupamiento.

Metas

- Desarrollo de herramientas de software de código abierto para la comunidad científica.
- Preparación y publicación de por lo menos un artículo en una revista indizada.
- Preparación y publicación de por lo menos un artículo en un congreso reconocido.
- Preparación de una tesis de licenciatura e iniciación de otras dos.
- Inducción de dos estudiantes a la ciencia computacional dentro del programa PROVERICyT 2016.

4. Metodología

Durante el presente año 2015 se han realizado las siguientes actividades preparativas para habilitar el proyecto propuesto y así asegurar que se pueda concluir dentro del periodo.

1. Se utilizó `wget` y un *script* de `bash` para descargar todos los boletos epidemiológicos disponibles.
2. Se extrajeron las páginas individuales de los archivos PDF como imágenes en formato PNG con `ImageMagick` (c.f. figura 1).
3. Se implementó un programa en `python` para convertir las imágenes extraídas en blanco y negro, eliminando los colores de relleno presentes en las cabeceras de los cuadros y también eliminando las líneas negras de los cuadros (la versión procesada de figura 1 se muestra en figura 2 en la página 5) para facilitar la fase de reconocimiento óptico de caracteres.
4. Se extrajo un listado de palabras que aparecen en el CIE-10 [9] y se agregaron manualmente palabras comúnmente utilizados en los boletines para preparar un diccionario customizado contra el cual se comparan los textos extraídos durante el reconocimiento óptico de caracteres.
5. Se procesó con `tesseract-ocr` a cada imagen para extraer en formato HOOCR los textos identificados, sus coordenadas y dimensiones.
6. Se implementó en `python` un programa que busca identificar si una página contiene un cuadro con datos de cada estado de la república, identificando la cabecera (que detalla las enfermedades incluidas en el cuadro) y acomodando el contenido en renglones (por estado de la república) y columnas (los conteos correspondientes a hombres y mujeres, totales y el acumulado correspondiente del año anterior).
7. Se identificaron a partir de datos públicos los centros poblacionales mayores de México³ y se obtuvo la población y las coordenadas geográficas de cada centro poblacional (latitud y longitud) con los datos de Wikipedia.
8. Se obtuvieron datos históricos climáticos [13] (temperatura y precipitación, mínimos y máximos diarios) para las tres estaciones de captura más cercanos a cada centro poblacional.

³Aquellas cuya población excede 100,000 habitantes; la población total de los centros incluidos supera a 51 millones.

Vigilancia Epidemiológica Semana 20, 2014

31

CUADRO 7.2 Casos por entidad federativa de Enfermedades Transmitidas por Vector hasta la semana epidemiológica 20 del 2014

ENTIDAD FEDERATIVA	Fiebre por Dengue CIE-10ª REV. A90					Fiebre Hemorrágica por Dengue CIE-10ª REV. A91				
	2014			2013		2014			2013	
	Confirmados			Confirmados	En Estudio	Confirmados			Confirmados	
	Sem.	Acum.	F	Acum.	Acum.	Sem.	Acum.	F	Acum.	
Aguascalientes	-	-	-	-	-	-	-	-	-	
Baja California	-	-	1	-	-	-	-	-	-	
Baja California Sur	6	67	58	46	6	2	8	1	1	
Campeche	6	50	56	51	5	-	10	8	19	
Coahuila	-	-	1	3	3	-	-	-	-	
Colima	2	49	69	269	26	-	28	21	92	
Chiapas	12	65	73	184	21	19	95	73	240	
Chihuahua	-	-	-	-	-	-	-	-	-	
Distrito Federal	-	-	-	-	-	-	-	-	-	
Durango	1	1	-	1	-	-	-	-	-	
Guanajuato	-	-	-	-	-	-	-	-	-	
Guerrero	17	130	105	473	71	9	74	74	296	
Hidalgo	1	5	2	13	-	-	-	-	1	
Jalisco	5	85	123	156	50	2	45	40	17	
México	-	-	-	9	3	-	-	-	-	
Michoacán	1	46	52	397	1	-	2	1	35	
Morelos	4	23	31	12	52	4	11	13	19	
Navarrit	6	21	37	53	173	3	48	50	121	
Nuevo León	46	157	329	413	1	-	1	2	2	
Oaxaca	5	49	68	176	40	7	50	46	83	
Puebla	8	29	26	157	-	-	-	-	1	
Querétaro	-	1	-	-	-	-	-	-	-	
Quintana Roo	3	51	49	284	62	-	39	37	213	
San Luis Potosí	-	24	28	157	-	-	1	1	7	
Sinaloa	2	115	135	68	477	6	151	143	19	
Sonora	7	43	40	23	17	-	16	13	1	
Tabasco	4	27	51	792	7	-	9	16	200	
Tamaulipas	4	19	19	101	8	-	2	4	37	
Tlaxcala	-	-	-	-	-	-	-	-	-	
Veracruz	2	63	92	328	56	1	53	42	276	
Yucatán	4	103	111	188	40	2	61	68	102	
Zacatecas	-	-	-	-	-	-	-	-	-	
TOTAL	146	1 223	1 556	4 354	1 119	55	704	653	1 782	

FUENTE: SINAVE/DGE/Salud 2014. Información preliminar.

Figura 2: El ejemplo de figura 1 procesado para eliminar rellenos de color y la reja del cuadro de datos numéricos.

- Se obtuvo un mapa de México del INEGI y se preparó un mapa en blanco y negro sin etiquetas de texto cubriendo todo México continental, separado a estados individuales con fronteras en negro y el interior en blanco. Se identificaron y almacenaron las posiciones en pixeles de cada estado para poder colorear un estado específico con un color específico; se muestra un mapa coloreado con datos al azar como ejemplo en figura 3 en la página 6.
- Se preparó un programa en python que toma como entrada series temporales de datos para los estados de la república y produce como salida (aprovechando las herramientas de ImageMagick) un mapa animado donde los estados vienen coloreados según el dato numérico asociado a cada uno a lo largo del periodo incluido en los datos.

La metodología por aplicar durante el año del proyecto propuesto, considerando los avances existentes, consiste en las siguientes fases:

- Entrenamiento de tesseract-ocr al tipo de letra de específica de los boletín y la restricción del diccionario a la versión específica construida a partir del CIE 10 para reducir errores de detección en la fase de reconocimiento óptico de caracteres.

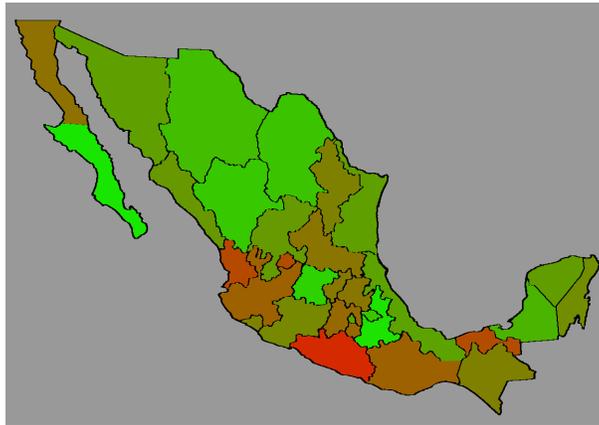


Figura 3: Un ejemplo de una visualización con mapa coloreado donde los colores de cada estado representan la magnitud de alguna variable numérica de interés; en este ejemplo los datos son simplemente números pseudoaleatorios, ya que aún no se cuenta con extracción precisa de series de tiempos (lo que se propone como parte de este proyecto). Estas mapas se pueden combinar en animaciones que muestran la evaluación de la serie de tiempo y se pueden colocar dos o más mapas juntos para visualizar múltiples series de tiempo de forma simultánea.

2. Extracción de archivos CSV (inglés: comma-separated values; son hojas de cálculo de texto plano) por cada cuadro, identificando las enfermedades reportadas en ello, la semana epidemiológica y el año en cuestión (con una combinación de *scripts* de *bash* y programas escritas en *python*).
3. Organización de los archivos CSV por enfermedad en forma de series de tiempo semanales por entidad federativa (se desarrolla una herramienta para automatizar esto).
4. Análisis estadístico de los datos climáticos de los centros poblacionales, agrupándolos por estado y cambiando la discretización a semanal (los datos crudos son diarios).
5. Combinación de la serie de tiempo de cada enfermedad con su respectivo serie de tiempo climático (se desarrolla una herramienta para automatizar esta tarea).
6. Visualización en *python* de las series de tiempo obtenidas con mapas coloreadas y gráficas de línea para las enfermedades y los datos climáticos con un intervalo de retraso ajustable entre el clima y la tasa de la enfermedad para inspección visual de los investigadores del área de salud para formular hipótesis sobre las dependencias entre las series y los efectos entre cercanía y ubicación geográfica de las entidades federativas.
7. Revisión detallada de estado de arte de modelos epidemiológicos que incorporan aspectos climáticos, temporales y espaciales.
8. Formulación de modelos a partir de las hipótesis que surgen del análisis preliminar; implementación y análisis de estos modelos para cada enfermedad; exploración de posibles factores relevantes adicionales, tales como frecuencia de tránsito vial entre centros poblacionales o índices de contaminación⁴
9. Caracterización de las distintas enfermedades según el comportamiento de su modelo; formación de hipótesis para investigación futura del área de la salud a base de los resultados observados; documentación detallada de los hallazgos en literatura científica; difusión de las herramientas desarrolladas.

⁴Contando con las series de tiempo de las enfermedades, se puede cruzar los datos con los indicadores socioeconómicos del INEGI, entre otras fuentes.

5. Infraestructura y apoyo técnico disponible

El cubículo del investigador responsable en el CIDICS cuenta con una computadora tipo estación de trabajo (facilitada en forma temporal) con 4 GB de memoria RAM (sistema operativo Ubuntu Linux 14.04 LTS), al cual se puede acceder vía SSH desde el cubículo del investigador responsable en la FIME a través de una iMac con 1 GB de RAM.

6. Formación de recursos humanos

Se contempla un trabajo de tesis de licenciatura concluida, dos trabajos de tesis iniciados, y dos estudiantes de verano científico dentro del programa PROVERICyT de la UANL⁵.

Estudiantes becarios

Los becarios del proyecto son tres estudiantes regulares de licenciatura del programa educativo *Ingeniero en Tecnología en Software* (ITS) de la FIME, UANL. Se anexan sus kárDEX y credenciales. Todos tienen promedios encima del mínimo requerido de 80 % y edades debajo del máximo permitido de 35 años.

a) *Jesús Alejandro Garza Leal*

- Matrícula: 1552425
- Semestre: 7°
- Promedio: 82
- Edad: 21

b) *Pamela González García*

- Matrícula: 1588003
- Semestre: 5°
- Promedio: 92
- Edad: 19

c) *Pedro Sánchez Martínez*

- Matrícula: 1590340
- Semestre: 5°
- Promedio: 85
- Edad: 19

Se contempla la tesis del estudiante de séptimo semestre como producto de este proyecto (referido como becario tesista en esta propuesta); las tesis de los estudiantes de quinto semestre se inician dentro del proyecto propuesto y se planea terminarlas como parte de un proyecto de continuación con fondos externos que se solicitará a base de los resultados de éste.

7. Calendarización trimestral de actividades

En la calendarización se estima que el proyecto inicie a medio semestre agosto-diciembre 2015 y termine un año después.

Primer trimestre (octubre-diciembre 2015). Adquisición de equipo y libros; introducción de los becarios a los prototipos previos desarrollados por el investigador responsable; desarrollo de una herramienta de software para la digitalización de cuadros de tasas de incidencia de los boletines epidemiológicos (en colaboración entre el investigador responsable y los estudiantes becarios).

Segundo trimestre (enero-marzo 2016). Desarrollo de una herramienta de software para la sincronización de los datos epidemiológicos con los climáticos y para la construcción de series de tiempo multifactoriales en colaboración entre el investigador responsable y los estudiantes becarios; inicio de la redacción de tesis.

Tercer trimestre (abril-junio 2016). Implementación de un modelo estadístico, adaptando técnicas existentes y herramientas computacionales de código abierto, en colaboración entre los investigadores participantes y los estudiantes becarios; redacción de artículos para revista y congreso; envío de artículo a congreso; finalización de la redacción de tesis.

⁵No se nombran los estudiantes PROVERICyT ya que su asignación no depende directamente del investigador responsable.

Cuarto trimestre (julio-septiembre 2016). Desarrollo de una herramienta de software para la clasificación de enfermedades; envío de artículo a revista; presentación del trabajo en un congreso (el becario tesista); divulgación de las herramientas desarrolladas; defensa de tesis de licenciatura; estancia de verano científico de dos alumnos PROVERICyT para análisis de datos; preparación de una propuesta de proyecto de ciencias básicas para continuar el trabajo el área de modelado y análisis epidemiológico.

8. Bibliografía

- [1] O. Adegboye and D. Kotze. Epidemiological analysis of spatially misaligned data: a case of highly pathogenic avian influenza virus outbreak in Nigeria. *Epidemiology and Infection*, 142(5):940–949, May 2014.
- [2] A. Costilla-Esquivel, F. Corona-Villavicencio, J. Velasco-Castañón, C. Medina-de la Garza, R. Martínez-Villarreal, D. Cortes-Hernández, L. Ramírez-López, and G. González-Farías. A relationship between acute respiratory illnesses and weather. *Epidemiology and Infection*, 142(7):1375–1383, July 2014.
- [3] M. Craig, R. Snow, and D. le Sueur. A climate-based distribution model of malaria transmission in sub-saharan africa. *Parasitology Today*, 15(3):105–111, Mar. 1999.
- [4] O. Diekmann and J. Heesterbeek. *Mathematical Epidemiology of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley, 2000. ISBN 978-0-471-49241-2.
- [5] Dirección General de Epidemiología. Plan nacional de preparación y respuesta ante la intensificación de la influenza estacional o ante una pandemia de influenza, 2010. URL http://www.epidemiologia.salud.gob.mx/doctos/infoepid/vig_epid_manuales/Plan_Nal_Pandemia_Influenza.pdf.
- [6] D. A. Ford, J. H. Kaufman, and I. Eiron. An extensible spatial and temporal epidemiological modelling system. *International Journal of Health Geographics*, 5(4), 2006. doi: 10.1186/1476-072X-5-4.
- [7] A. C. Gatrell, T. C. Bailey, P. J. Diggle, and B. S. Rowlingson. Spatial point pattern analysis and its application in geographical epidemiology. *Transactions of the Institute of British Geographers*, 21(1):256–274, 1996.
- [8] K. Jalava, J. Sanea, J. Ollgren, R. Ruuhela, O. Rätti, S. Kurkela, P. Helle, S. Hartonen, P. Pirinen, O. Vapalahti, and M. Kuusi. Climatic, ecological and socioeconomic factors as predictors of Sindbis virus infections in Finland. *Epidemiology and Infection*, 141(9):1857–1866, Sept. 2013.
- [9] Organización Mundial de la Salud. Clasificación internacional de enfermedades 1º, May 1990. 43rd World Health Assembly.
- [10] D. U. Pfeiffer, T. P. Robinson, M. Stevenson, K. B. Stevens, D. J. Rogers, and A. C. A. Clements. *Spatial Analysis in Epidemiology*. Oxford University Press, 2008. ISBN 978-0198509899.
- [11] L. A. Rvachev and I. M. Longini Jr. A mathematical model for the global spread of influenza. *Mathematical Biosciences*, 75(1):3–22, July 1985.
- [12] S. Selvin. *Statistical Analysis of Epidemiologic Data*. Monographs in Epidemiology and Biostatistics. Oxford University Press, 3 edition, 2004. ISBN 978-0195172805.
- [13] E. Valenzuela and T. Cavazos. CLICOM — base de datos climatológica nacional. CICESE, 2013. URL <http://clicom-mex.cicese.mx/>.
- [14] J. G. Velasco-Castañón. México ante el mundo: la reciente pandemia de Influenza. *Medicina Universitaria*, 12(49):248–9, 2010.

9. Participantes

Responsable

- a) Nombre: *Satu Elisa Schaeffer*
- b) Grado académico: *Doctor en Ciencias en Tecnología en Ciencia e Ingeniería de la Computación*
- c) Nivel en el SNI: **1** vigencia 2018
- d) Últimas 5 publicaciones:
 - i) S.E. Garza V. & S.E. Schaeffer. **Local bilateral clustering for identifying research topics and groups from bibliographical data.** *Knowledge and Information Systems*, 2015. DOI 10.1007/s10115-015-0867-y, ISSN 0219-1377.
 - ii) C.A. Castillo S., A. Conde E. & S.E. Schaeffer. **Directional Overcurrent Relay Coordination Considering Non-Standardized Time Curves.** *Electric Power Systems Research* 122:42â€”49, 2015. ISSN 0378-7796.
 - iii) S.C. Urbina C., S.E. Schaeffer & S.E. Garza V. **A visual aid on the ecological impact of products for online retail.** *Research in Computing Science* 81:103â€”112, 2014. ISSN 1870-4069.
 - iv) A. Tchernykh, J.E. Pecero, A. Barrondo & S.E. Schaeffer. **Adaptive Energy Efficient Scheduling in Peer-to-Peer Desktop Grids.** *Future Generation Computer Systems* 36:209â€”220, 2014. ISSN 0167-739X.
 - v) J.C. Espinosa C., S.E. Schaeffer & S.E. Garza V. **Augmented reality for green consumption: using computer vision to inform the consumers at time of purchase.** In A. Gelbukh, F. Castro Espinoza, and S.N. Galicia-Haro, editors, *Proc. of 2014 Thirteenth Mexican International Conference on Artificial Intelligence – Special Session*, pp. 45–51, IEEE Computer Society. ISBN 978-1-4799-9900-2.

Colaboradores

- a) Nombre: *José Gerardo Velasco Castañón*
- b) Grado académico: *Doctor en Ciencias en Microbiología*
- c) Nivel en el SNI: N/A
- d) Últimas 5 publicaciones:
 - i) J.G. Velasco C. & C.E. Medina de la G. **Fear of the unknown: Influenza vaccination.** *Medicina Universitaria* 16:87-9, 2014.
 - ii) A. Costilla E., F. Corona V., J.G. Velasco C., C.E. Medina de la G., R.T. Martínez V., D.E. Cortes H., L.E. Ramírez L. & G. González F. **A relationship between acute respiratory illnesses and weather.** *Epidemiology and Infection* 142(7): 1375–1382, 2014.
 - iii) J.G. Velasco C. **México ante el mundo: la reciente pandemia de Influenza.** *Medicina Universitaria* 12(49):248-9, 2010.
 - iv) J.G. Velasco C. **El affaire A H1N1: de un brote epidémico a pandemia.** *Medicina Universitaria* 11(43):87-88, 2009.
 - v) J.G. Velasco C, G.I. Martínez T., M.G. Rodríguez L., J.C. Argüello E. & L.E. Ramírez L. **Observaciones sobre el desarrollo del dengue en Nuevo León.** *Medicina Universitaria* 8(31):156-8, 2006.
- e) Participación en redacción de artículos, interpretación de resultados obtenidos & aplicación de modelos epidemiológicos.

- a) Nombre: *Moisés Hinojosa Rivera*
- b) Grado académico: *Doctor en Ciencias de la Ingeniería Mecánica con especialidad en Materiales*
- c) Nivel en el SNI: **1**, vigencia 2017
- d) Últimas 5 publicaciones:
- i) D. Aguirre, J. Aldaco & M. Hinojosa R. **Microestructura, propagación y emisión acústica de grietas en dos papeles de celulosa**. *Revista Mexicana de Física* 61:254–260, 2015.
 - ii) L. Chávez G., B. Medina L., R.F. Cienfuegos, M.A. Garza N, R.N. Vannier, A. Ringuedé, M. Hinojosa R. & M. Cassir: **Synthesis and characterization of LaNi_{1-x}Co_xO₃: Role of microstructure on magnetic properties**. *Journal of Rare Earths* 33(3), 2015.
 - iii) J. Aldaco & M. Hinojosa R. **Simulación atomística de la fractura dinámica en un material frágil**. *Revista Mexicana de Física* 58(6):466–471, 2012.
 - iv) L. Chávez G. & M. Hinojosa R. **Bagasse from the mezcal industry as an alternative renewable energy produced in arid lands**. *Fuel* 89(12):4049–4052, 2010.
 - v) L. Chávez G., F.J. Garza & M. Hinojosa R. **Scaling exponents for fracture surfaces in opal glass**. *Materials Science and Engineering A* 527(24):6474–6477, 2010.
- e) Participación en redacción de artículos, formulación de modelos matemáticos de propagación, interpretación de resultados obtenidos & integrante del comité de tesis de los becarios.



- a) Nombre: *Sara Elena Garza Villarreal*
- b) Grado académico: *Doctor en Tecnologías de Información y Comunicaciones con Especialidad en Sistemas Inteligentes*
- c) Nivel en el SNI: **C**, vigencia 2016
- d) Últimas 5 publicaciones:
- i) S.E. Garza V. & S.E. Schaeffer. **Local bilateral clustering for identifying research topics and groups from bibliographical data**. *Knowledge and Information Systems*, aceptado, 2015. DOI 10.1007/s10115-015-0867-y, ISSN 0219-1377.
 - ii) S.C. Urbina C., S.E. Schaeffer & S.E. Garza V. **A visual aid on the ecological impact of products for online retail**. *Research in Computing Science* 81:103–112, 2014. ISSN 1870-4069.
 - iii) J.C. Espinosa C., S.E. Schaeffer & S.E. Garza V. **Augmented reality for green consumption: using computer vision to inform the consumers at time of purchase**. In A. Gelbukh, F. Castro Espinoza, and S.N. Galicia-Haro, editors, *Proc. of 2014 Thirteenth Mexican International Conference on Artificial Intelligence – Special Session*, pp. 45–51, IEEE Computer Society. ISBN 978-1-4799-9900-2.
 - iv) J.A. Cordero C., S.E. Garza V. & S.E. Schaeffer. **Entity Recognition for Duplicate Filtering**. En Traina, et al. (eds), *Similarity Search and Applications*, LNCS 8821, pp. 253–264, Springer, 2014.
 - v) S.E. Garza V. & F. Torres G. **Expertise Assessment: A Quantitative Approach Using Natural Semantic Networks**. *Daena: International Journal of Good Conscience* 9:76–85, 2014.
- e) Participación en redacción de artículos, agrupamiento y análisis de datos & integrante del comité de tesis de los becarios.

10. Justificación financiera de los requerimientos

Se detalla lo solicitado en cada partida.

SUMA TOTAL \$70,000 M.N.

Rubro 100: Gasto Corriente

- 101 Viáticos** Apoyo para la presentación de los resultados del proyecto: alimentación y hospedaje del estudiante al participar en un congreso para presentar su trabajo de tesis, \$ 3,500 M.N.; se respetarán los límites establecidos según el destino.
- 102 Pasajes** Pago de transporte del becario tesista a un congreso para presentar resultados de su trabajo de tesis: \$3,500 M.N. El gasto ejercido entre los rubros de viáticos y pasajes corresponde al 10 % del monto apoyado, lo que es el máximo permitido.
- 103 Gastos de Trabajo de Campo** No se requiere este tipo de apoyo.
- 104 Ediciones e Impresiones** Pagos de los trabajos de impresión a color y encuadernación de tesis (solamente se cuenta con impresora blanco-y-negro en la oficina), \$4,000 M.N.
- 105 Servicios Externos** No se requiere pagos a terceros.
- 106 Cuotas de Inscripción** (a) Cuota de afiliación a sociedad científica ACM (la actual vence el 31 de octubre del 2015), incluyendo acceso al repositorio digital de la asociación, por un año: USD 1,000 ≈ \$1,635 M.N.; (b) pago para la asistencia a congreso para que el becario tesista presente su trabajo, \$2,000 M.N.
- 107 Artículos, Materiales y Útiles Diversos** Compra de materiales para actividades diarias: papelería de oficina, consumibles, tóner de impresora: \$1,867 M.N.
- 108 Libros y Revistas Técnicas y Científicas** Se contempla adquisición de libros según las necesidades de los alumnos becarios por un monto total de \$3,500 M.N.
- 109 Animales para Rancho y Granja** No se requiere compra de animales.
- 110 Becas** Apoyo para un alumno colaborador para concluir su tesis de licenciatura. El alumno becario propuesto no cuenta con otra beca; se anexan kárdex vigente y copia de una identificación oficial. Son tres becas de \$5,000 (dividido en dos pagos semestrales) y se incluyen dos becas PROVERICyT de \$1,000 cada una. El total de becas corresponde al 24 % del monto apoyado al proyecto, lo que es dentro de lo permitido.

Se resumen los montos por rubro, indicando también el porcentaje del presupuesto total del proyecto.

CLAVE	PARTIDA	IMPORTE	PORCENTAJE
101	Viáticos	\$3,500	5 %
102	Pasajes	\$3,500	5 %
103	Gastos de Trabajo de Campo	<i>No aplica</i>	0 %
104	Ediciones e Impresiones	\$5,000	5 %
105	Servicios Externos	<i>No aplica</i>	0 %
106	Cuotas de Inscripción	\$2,635	4 %
107	Artículos, Materiales y Útiles Diversos	\$1,867	3 %
108	Libros y Revistas Técnicas y Científicas	\$3,500	5 %
109	Animales para Rancho y Granja	<i>No aplica</i>	0 %
110	Becas	\$17,000	24 %

TOTAL DE GASTO CORRIENTE \$36,002 (51 %)

Rubro 200: Gasto de Inversión

Se anexan cotizaciones realizadas en línea con vendedores nacionales estables, incluyendo IVA en ambos casos.

201 Equipo de Apoyo y Laboratorio No se requiere compra de equipo de este tipo.

202 Equipo de Computación La computadora principal de la oficina de Dra. Schaeffer es una iMac que fue comprado con el PAICyT 2007 (CA1475-07) y por su edad sus especificaciones ya no son suficientes para software actual (tiene solamente 1 GB de memoria RAM), por lo cual se solicita para los experimentos computacionales del trabajo de tesis una versión moderna del mismo tipo, iMac de 27 pulgadas, con un procesador de cuatro núcleos para el procesamiento paralelo requerido y 8 GB de RAM, para el uso de los estudiantes becarios, con un costo cotizado de \$31,999 M.N. Además se requiere es una impresora-escáner (fija, no portátil), debido a que el anterior que se tuvo descompuso en el 2015 (se ha entregado el aparato descompuesto al departamento de registro patrimonial para la baja correspondiente), con un costo cotizado de \$1,999 M.N.

CLAVE	PARTIDA	IMPORTE	PORCENTAJE
201	Equipo de Apoyo y Laboratorio	No aplica	0 %
202	Equipo de Computación	\$33,998	49 %

TOTAL DE GASTO DE INVERSIÓN \$33,998 M.N. (49 %)

Evidencia de artículo del proyecto anterior

Se ha preparado la propuesta acorde al manual de operaciones⁶. Esta última sección es referente a la convocatoria 2015⁷ que además instruye al proponenta a especificar evidencia de un producto del proyecto financiado en PAICyT 2012.

A partir del trabajo relacionado al proyecto PAICyT anterior de la Dra. Schaeffer (IT553-10; *Sistemas inteligentes en apoyo a la toma de decisiones: Inteligencia de negocios y el manejo de datos personales legaliforme en las PyME*), se ha enviado a la revista *Decision Support Systems* un manuscrito titulado *Static R&D project portfolio selection in public organizations* con número de manuscrito DEC SUP-D-14-00205, en coautoría con Dr. Fernando López (profesor de la FIME, UANL), M.C. Nancy Arratia (estudiante de la FIME, UANL) y Dra. Laura Cruz (profesora del ITCM). El manuscrito se encuentra en el proceso de revisión desde marzo del 2014; se enviarán las últimas correcciones al editorial en este mes de agosto del 2015 y se espera recibir la respuesta antes del fin del año (el revisor #1 ya encuentra satisfecho con el manuscrito y se realizaron los ajustes solicitados por el revisor #2). El editor responsable es el editor principal de la revista, James R. Marsden. La evidencia anexada es la segunda revisión, para la cual estamos por terminar las correcciones solicitadas por lo revisores.

⁶<http://www.uanl.mx/sites/default/files/documentos/universidad/manual-de-operaciones-paicyt.pdf>

⁷<http://www.uanl.mx/sites/default/files/documentos/convocatoria/3741/convocatoria-paicyt-2015.pdf>