



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"
"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

SAN LUIS, 14 AGO. 2017

VISTO:

El Expediente EXP-USL: 7691/2017 mediante el cual se solicita la protocolización del Curso de Posgrado: **NUEVAS PROPUESTAS PARA BÚSQUEDAS POR SIMILITUD EN BASES DE DATOS MÉTRICAS**; y

CONSIDERANDO:

Que el Curso de Posgrado se dictará en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales del 8 al 18 de agosto de 2017, con un crédito horario de 60 horas presenciales y bajo la coordinación de la Dra. Nora **REYES**.

Que la Comisión Asesora de Posgrado de la Facultad de Ciencias Físico Matemáticas y Naturales recomienda aprobar el curso de referencia.

Que el Consejo de Posgrado de la Universidad Nacional de San Luis en su reunión del 24 de julio de 2017, analizó la propuesta y observa que el programa del curso, bibliografía, metodología de evaluación y docentes a cargo, constituyen una propuesta de formación de posgrado de calidad en su campo específico de estudio.

Que, por lo expuesto, el Consejo de Posgrado aprueba la propuesta como Curso de Posgrado, según lo establecido en Ordenanza CS N° 35/16.

Que corresponde su protocolización.

Por ello y en uso de sus atribuciones

EL RECTOR DE LA UNIVERSIDAD NACIONAL DE SAN LUIS

RESUELVE:

ARTÍCULO 1°.- Protocolizar el dictado del Curso de Posgrado: **NUEVAS PROPUESTAS PARA BÚSQUEDAS POR SIMILITUD EN BASES DE DATOS MÉTRICAS**, en el ámbito de la Facultad de Ciencias Físico Matemáticas y Naturales del 8 al 18 de agosto de 2017, con un crédito horario de 60 horas presenciales.

ARTÍCULO 2°.- Protocolizar el cuerpo docente constituido por: Responsables: Dra. Karina **FIGUEROA MORA** (Pas. G24678382) de la Universidad Michoacana de San Nicolás de Hidalgo – México, Dra. Nora **REYES** (DNI N° 16.865.387),

Cpde RESOLUCIÓN R N° **1269**

Dr. Jose Roberto Saad
Vice Rector - UNSL
A/C Rectorado RR N

Oscar Guillermo Segura
Director de Despacho
UNSL



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"

"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL


Auxiliares: Lic. Verónica del Rosario **LUDUEÑA** (DNI N° 17.077.521), Mg. Patricia **ROGGERO** (DNI N° 17.779.819) todas de esta Casa de Estudios Superiores.

ARTÍCULO 3°.- Aprobar el programa del Curso de referencia, de acuerdo al **ANEXO** de la presente disposición.-

ARTÍCULO 4°.- Comuníquese, insértese en el Libro de Resoluciones, publíquese en el Digesto Electrónico de la UNSL y archívese.-

RESOLUCIÓN R N° 1269
mav


Verónica del Rosario LUDUEÑA
ACCIONARIA DE ESTUDIOS SUPERIORES
UNSL


Dr. José Roberto Saad
Vicerrector - UNSL
A/C Rectorado RR N 1421/17



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"

"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

ANEXO

IDENTIFICACIÓN DEL CURSO

UNIDAD ACADÉMICA RESPONSABLE: Facultad de Ciencias Físico Matemáticas y Naturales

DENOMINACIÓN DEL CURSO: NUEVAS PROPUESTAS PARA BÚSQUEDAS POR SIMILITUD EN BASES DE DATOS MÉTRICAS

CATEGORIZACIÓN: Perfeccionamiento

FECHA DE DICTADO DEL CURSO: 8 al 18 de agosto de 2017

MODALIDAD DE DICTADO: Presencial

CRÉDITO HORARIO TOTAL: 60 horas (20 hs. teóricas, 20 hs. de prácticas de Aula y 20 hs. de prácticas de Laboratorio)

COORDINADORA: Dra. Nora REYES (DNI N° 16.865.387)

EQUIPO DOCENTE

RESPONSABLES: Dra. Karina FIGUEROA MORA, Dra. Nora REYES

AUXILIARES: Lic. Verónica del Rosario LUDUEÑA, Mg. Patricia ROGGERO

PROGRAMA ANALÍTICO

FUNDAMENTACIÓN: En la actualidad los Sistemas Administradores de Bases de Datos (DBMS) incorporan la capacidad de almacenar nuevos tipos de datos tales como imágenes u "objetos multimedia" y es cada vez más evidente la necesidad de procesar grandes conjuntos de datos, de manera tal de poder obtener información útil a partir de ellos. Sin embargo, en estas bases de datos, el concepto de búsqueda que puede aplicarse es el de *búsqueda por similitud*, es decir buscar elementos de la base de datos que sean similares o "próximos" a un elemento de consulta dado. Como la (di)similitud suele medirse con una métrica, a esta clase de base de datos se lo denomina *métrica*.

Por lo tanto, los profesionales de las Ciencias Informáticas necesitan perfeccionarse en esta área y conocer Algoritmos y Estructuras de Datos que se utilizan para Bases de Datos Métricas, para contar con las herramientas que le permitan resolver problemas que constituyen desafíos en la actualidad.

Este curso se enmarca dentro de las actividades conjuntas que se vienen desarrollando entre la Universidad Nacional de San Luis, la Universidad Michoacana de San Nicolás de Hidalgo (México) y la Universidad de Talca (Chile), en esta área.

OBJETIVOS:

El objetivo del curso es potenciar la investigación en esta área, mostrando los fundamentos y avances en los aspectos algorítmicos y de estructuras de datos de aplicaciones de búsqueda por similitud.

Cpde RESOLUCIÓN R N° 1269

[Handwritten signature]
Dr. Oscar Guillermo Segura
Vicepresidente - UNSL
A/C Rectorado RR N

[Handwritten signature]
Dr. Oscar Guillermo Segura
A/C Rectorado - UNSL



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"

"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

Al finalizar el curso se espera que el alumno conozca herramientas y pueda desarrollar también algunas propias para poder resolver diversos problemas de búsqueda por similitud para bases de datos métricas.

CONTENIDOS MÍNIMOS:

1. Conceptos Fundamentales de Espacios Métricos
2. Índices para Bases de Datos Métricas
3. Algoritmos Exactos y Aproximados
4. Otras operaciones de Interés sobre Bases de Datos Métricas

PROGRAMA DETALLADO:

1. Conceptos Fundamentales de Espacios Métricos:
 - a. Definición de espacios métricos.
 - b. Funciones de Distancia: propiedades.
 - c. Tipos de búsquedas por similitud más comunes.
 - d. Maldición de la dimensión.
 - e. Medidas de evaluación de la dimensionalidad.
2. Índices para Bases de Datos Métricas
 - a. Taxonomía de los índices.
 - b. Principales referentes de índices basados en particiones compactas.
 - c. Principales referentes de índices basados en pivotes.
 - d. Índices estáticos y dinámicos. Ejemplos.
 - e. Índices para memoria secundaria. Ejemplos.
3. Algoritmos Exactos y Aproximados
 - a. Algoritmos Exactos.
 - b. Algoritmos Aproximados.
 - c. Índices Basados en Permutaciones (exactos y aproximados).
 - d. Medidas de evaluación de calidad de respuesta.
4. Otras operaciones de Interés sobre Bases de Datos Métricas:
 - a. Join por Similitud, variantes.
 - b. Algoritmos para Join: con índices y sin índices.
 - c. Ejemplos de soluciones existentes.

SISTEMA DE EVALUACIÓN:

Para aprobar el curso el alumno deberá realizar los prácticos de aula y de laboratorio, y además deberá aprobar la evaluación final en la modalidad que el alumno elija. La evaluación final puede consistir en realizar un trabajo final o rendir un examen final.

En caso de elegir el trabajo final, el alumno podrá optar entre las siguientes opciones:

1. Realizar un trabajo de investigación acotado sobre un tema relacionado al curso.
2. Analizar y luego realizar una presentación oral, a los demás alumnos del curso, de un artículo de interés sobre los temas del curso.
3. Realizar una propuesta de mejora a alguno de los índices existentes para bases de datos métricas.

La evaluación será de carácter individual.

Cpde RESOLUCIÓN R N° 1269


Dr. José Roberto Szad
Vicerrector - UNSL
A/C Rectorado RR


Oscar Guillermo Segura
Director de Despacho
UNSL



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"

"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

BIBLIOGRAFÍA:

- * Hanan Samet. "Foundations of Multidimensional and Metric Data Structures". Morgan Kaufmann Publishers Inc., 2006.
- * Pavel Zezula, Giuseppe Amato, Vlastislav Dohnal, and Michal Batko. "Similarity Search: The Metric Space Approach (Advances in Database Systems)". Springer- Verlag New York, Inc., Secaucus, NJ, USA, 2005.
- * Edgar Chávez, Gonzalo Navarro, Ricardo Baeza-Yates and José Luis Marroquín, "Searching in Metric Spaces", ACM Computing Surveys, Vol. 33, No. 3, September 2001, pp. 273–321.
- * Jeffrey Scott Vitter, "External memory algorithms and data structures: dealing with massive data", ACM Computing Surveys, Vol. 33, Number 2, 2001. ISSN 0360-0300, Pages 209-271, doi = <http://doi.acm.org/10.1145/384192.384193>. ACM Press.
- * Ricardo Baeza-Yates and Berhier Ribeiro-Neto. Modern Information Retrieval: The Concepts and Technology Behind Search. Addison Wesley, 2011.
- * Edgar Chávez and Gonzalo Navarro. "A compact space decomposition for effective metric indexing". Pattern Recognition Letters 26, 9, 1363–1376. 2005.
- * Paolo Ciaccia, Marco Patella, Pavel Zezula. "M-tree: An efficient access method for similarity search in metric spaces". In Proc. of the 23rd Conference on Very Large Databases (VLDB'97). 426–435. 1997.
- * Gisli Hjaltason and Hanan Samet. "Index-driven similarity search in metric spaces". ACM Trans. on Database Systems 28, 4, 517–580. 2003.
- * Christian Böhm, Stefan Berchtold, S. and Daniel Keim. "Searching in high-dimensional spaces: Index structures for improving the performance of multimedia databases". ACM Computing Surveys 33, 3 (Sept.), 322–373. 2001.
- * Edgar Chavez, Karina Figueroa and Gonzalo Navarro. "Effective Proximity Retrieval by Ordering Permutations". IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). Vol. 30 No. 9. pp 1647-1658, sept 2008.
- * Veronica Gil-Costa, Mauricio Marin and Nora Reyes. "Parallel Query Processing on Distributed Clustering Indexes". Journal of Discrete Algorithms (7) 03-17, March 2009 (Elsevier).
- * Rodrigo Paredes and Nora Reyes. "Solving similarity joins and range queries in metric spaces with the list of twin clusters", Journal of Discrete Algorithms, Elsevier. Volume 7, Issue 1, Pages 18-35. ISSN 1570-8667, March 2009.
- * Gonzalo Navarro, Nora Reyes. "Dynamic Spatial Approximation Trees", en The ACM Journal of Experimental Algorithmics, Vol. 12, June 2008, SECTION: 1 - Regular Papers, Artículo 1.5, Pág. 1-68. ISSN 1084-6654. 2008.
- * Lieberman, M.D. and Sankaranarayanan, J. and Samet, H.. "A Fast Similarity Join Algorithm Using Graphics Processing Units". ICDE 2008. IEEE 24th International Conference on Data Engineering 2008 - Pp 1111:1120 - April 2008.
- * Christian Böhm, Bernhard Braunmüller, Florian Krebs, and Hans-Peter Kriegel. Epsilon grid order: an algorithm for the similarity join on massive high-dimensional data. In Proceedings of the 2001 ACM SIGMOD international conference on Management of data, SIGMOD '01, pages 379–388, New York, NY, USA, 2001. ACM.
- * Edgar Chávez, Verónica Ludeña, Nora Reyes, and Patricia Roggero. Faster proximity searching with the distal sat. Available online 9 January 2016. Vol. 59, July 2016, Pages 15–47. ISSN 0306-4379.

[Handwritten signature]
Dr. José Roberto Saad
Vicerrector F - UNSL
A/C Rectorado RR

[Handwritten signature]
Dr. Jorge Armando Ferrer
A/C Rectorado RR - UNSL



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"

"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

- * Edgar Chávez, José L. Marroquín, and Gonzalo Navarro. Fixed queries array: A fast and economical data structure for proximity searching. *Multimedia Tools and Applications*, 14(2):113–135, 2001.
- * Edgar Chávez and Gonzalo Navarro. A compact space decomposition for effective metric indexing. *Pattern Recognition Letters*, 26(9):1363–1376, 2005.
- * Paolo Ciaccia and Marco Patella. Approximate and probabilistic methods. *SIGSPATIAL Special*, 2(2):16–19, July 2010.
- * Luiz Olmes Carvalho, Lucio F. D. Santos, Willian D. Oliveira, Agma Juci Machado Traina, and Jr. Traina, Caetano. Similarity joins and beyond: An extended set of binary operators with order. In *Similarity Search and Applications*, volume 9371 of *Lecture Notes in Computer Science*, pages 29–41. Springer International Publishing, 2015.
- * Michel Marie Deza and Elena Deza. *Encyclopedia of Distances*. Springer Berlin Heidelberg, 2009.
- * Kimmo Fredriksson and Billy Braithwaite. Quicker range- and k-nn joins in metric spaces. *Information Systems*, 52:189 – 204, 2015. Special Issue on Selected Papers from SISAP 2013.
- * Edwin H. Jacox and Hanan Samet. Metric space similarity joins. *ACM Trans. Database Syst.*, 33(2), 2008.
- * Gonzalo Navarro, Rodrigo Paredes, Nora Reyes, and Cristian Bustos. An empirical evaluation of intrinsic dimension estimators. Available online 16 June 2016. Vol. 64, March 2017, Pages 206–218. ISSN 0306-4379.
- * Gonzalo Navarro and Nora Reyes. Dynamic spatial approximation trees. *Journal of Experimental Algorithmics*, 12:1.5:1–1.5:68, June 2008. ISSN 1084-6654.
- * Gonzalo Navarro and Nora Reyes. New dynamic metric indices for secondary memory. Available online 7 April 2016. Vol. 59, July 2016, Pages 48–78. ISSN 0306-4379.
- * Karina Figueroa, Rodrigo Paredes, J. Antonio Camarena-Ibarrola, and Nora Reyes. Proceedings of the 9th Mexican Conference on Pattern Recognition (MCPR2017), *Lecture Notes in Computer Science series*, Volume 10267. Pages 74 – 83, 2017. Springer 2017. ISBN 978-3-319-59225-1 (Print) ISBN 978-3-319-59226-8 (eBook).
- * Lucio F. D. Santos, Luiz Olmes Carvalho, Willian D. Oliveira, Agma J.M. Traina, and Jr. Traina, Caetano. Diversity in similarity joins. In *Similarity Search and Applications*, volume 9371 of *Lecture Notes in Computer Science*, pages 42–53. Springer International Publishing, 2015.
- * Yasin N. Silva, Spencer S. Pearson, Jaime Chon, and Ryan Roberts. Similarity joins: Their implementation and interactions with other database operators. *Information Systems*, 52:149 – 162, 2015. Special Issue on Selected Papers from SISAP 2013.
- * Karina Figueroa and Rodrigo Paredes. Boosting the Permutations based Index for Proximity Searching. *MCPR 2015. LNCS*. Springer. ISSN: 1870-406. pag 103-112. ISBN: 978-3-319-19263-5.
- * Karina Figueroa, Rodrigo Paredes. Compact and Efficient Permutations for Proximity Searching. *MCPR 2012. LNCS 7329*. pp 207–215. ISSN 0302-9743.
- * Karina Figueroa, Rodrigo Paredes. Dynamic Permutation Based Index for Proximity Searching. *SISAP 2015, LCNS 8199*, Springer.
- * Karina Figueroa, Rodrigo Paredes. An Efficient Permutant Selection Heuristic for Proximity Searching in Metric Spaces. *MCPR 2014. LNCS*. Vol 8495 . Springer. ISSN: 0302-9743. pag 102-111
- * Karina Figueroa, Rodrigo Paredes. List of Clustered Permutations for Proximity Searching. *SISAP LNCS 8199*, pp 50. Springer

[Handwritten signature]
Dr. Oscar Roberto Saad
Vicepresidente - UNSL
A/C Rectorado RR N°

[Handwritten signature]
Dr. Oscar Roberto Saad
A/C Rectorado - UNSL

Cpde RESOLUCIÓN R N° 1269



Universidad Nacional de San Luis
Rectorado

"2017 – AÑO DE LAS ENERGÍAS RENOVABLES"

"Centenario del Natalicio del Poeta Puntano Antonio Esteban AGÜERO"

ES COPIA
OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL

CARACTERÍSTICAS DEL CURSO

DESTINATARIOS Y REQUISITOS DE INSCRIPCIÓN: Egresados con título de grado universitario de 4 años o más en Ciencias Informáticas y en disciplinas afines a la temática del curso.

CUPO: máximo 20 personas.

PROCESO DE ADMISIÓN: En caso que el número de inscriptos supere el cupo, el cuerpo docente dará prioridad al orden de inscripción.

CRONOGRAMA DE ACTIVIDADES: las actividades iniciarán el 8 de agosto y se acordará con los alumnos admitidos los restantes horarios de clase para las clases teórico-prácticas y las consultas para las prácticas de laboratorio.

LUGAR DE DICTADO: Sala de Posgrado II del Departamento de Informática.

FECHA PREVISTA PARA ELEVAR LA NÓMINA DE ALUMNOS APROBADOS: Diciembre de 2017

FINANCIAMIENTO DEL CURSO

COSTOS: Los costos corresponden al traslado y estadía de la docente extranjera y los relativos a la dedicación horaria de los docentes locales.

FUENTES DE FINANCIAMIENTO: Universidad Michoacana de San Nicolás de Hidalgo (México), UNSL y aporte de los aranceles.

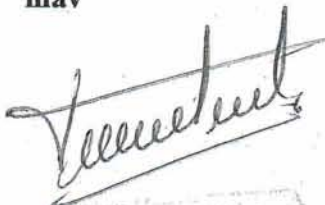
ARANCEL GENERAL: \$1.000 (pesos un mil).


BECA AL DOCENTE DE LA UNSL: Beca del 100%

BECA AL ALUMNO DE LA UNSL: Beca del 100%

OTRAS BECAS: El cuerpo docente podrá otorgar otras becas si lo considera pertinente.

Cpde RESOLUCIÓN R N° 1269
may


OSCAR GUILLERMO SEGURA
Director de Despacho
UNSL


Dr. José Roberto Saad
Vicedirector - UNSL
A/C Rectorado RR N 141/17