

Dr. David Eduardo Pinto Avendaño

Adscripción: Facultad de Ciencias de la Computación, BUAP, Puebla, México

Teléfono: (+52-222) 2295500 Ext. 2856

Correo electrónico y página Web: dpinto@cs.buap.mx / <http://dpinto.cs.buap.mx>

Membresías:

- SNI Nivel 1 (2013-2015), Area IV: Humanidades y Ciencias de la Conducta: Lingüística Computacional
- Perfil PROMEP
- Miembro de la Red Tecnologías del Lenguaje de CONACYT
- Sociedad Mexicana del procesamiento del lenguaje natural
- Sociedad Española del procesamiento del lenguaje natural
- Asociación Mexicana de visión computacional, cómputo neuronal y robótica

Formación académica:

- Licenciatura en Ciencias de la Computación, B. Universidad Autónoma de Puebla, 27/06/1997
- Maestría en Ciencias de la Computación, B. Universidad Autónoma de Puebla, 18/06/1999
- Doctorado en Informática, Universidad Politécnica de Valencia, España, 15/07/2008

Publicaciones (desde el año 2005):

Revistas:

(JCR/CONACYT)

- Grigori Sidorov, Alexander F. Gelbukh, Helena Gómez-Adorno, David Pinto: Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model. *Computación y Sistemas* 18(3) (2014)
- Fernando Perez-Tellez, John Cardiff, Paolo Rosso, David Pinto: Weblog and short text feature extraction and impact on categorisation. *Journal of Intelligent and Fuzzy Systems* 27(5): 2529-2544 (2014)
- David Pinto, Helena Gómez-Adorno, Darnes Vilariño Ayala, Vivek Kumar Singh: A graph-based multi-level linguistic representation for document understanding. *Pattern Recognition Letters* 41: 93-102 (2014)
- Paolo Rosso, Marcelo Errecalde, David Pinto: Analysis of short texts on the Web: Introduction to special issue. DOI: 10.1007/s10579-013-9220-9. *Language Resources and Evaluation* 47(1): 123-126 (2013)
- José A. Reyes, Azucena Montes, Juan G. González, David Eduardo Pinto: Clasificación de roles semánticos usando características sintácticas, semánticas y contextuales. *Computación y Sistemas* 17(2) (2013)
- David Pinto, Darnes Vilariño, Carlos Balderas, Mireya Tovar and Beatriz Beltran, Evaluating n-gram Models for a Bilingual Word Sense Disambiguation Task, *Computación y Sistemas* 15(2), 2011
- Claudia Denicia-Carral, Manuel Montes-y-Gómez, Luis Villaseñor-Pineda and David Pinto-Avendaño, Bilingual Document Clustering: Evaluating Cognates as Features, *Journal of Information and Library Science*, 35(3): 265-286, 2011
- David Pinto, Paolo Rosso, Hector Jiménez-Salazar: A Self-Enriching Methodology for Clustering Narrow Domain Short Texts, DOI:10.1093/comjnl/bxq069, *Computer Journal* 54(7): 1148-1165 (2011) 2010
- David Pinto, Jorge Civera, Alberto Barrón-Cedeño, Alfons Juan, Paolo Rosso: A statistical approach to crosslingual natural language tasks. DOI: 10.1016/j.jalgor.2009.02.005 <http://dx.doi.org/10.1016/j.jalgor.2009.02.005>, *Journal of Algorithms in Cognition, Informatics and Logic*. 64(1): 51-60, 2009

(Latin Index)

- Yuridiana Alemán, Darnes Vilariño, David Pinto, Mireya Tovar. Detección de depredadores sexuales utilizando un sistema de consulta y clasificación supervisada. *Research in Computing Science*. Vol 72. ISSN: 1870-4069. Instituto Politécnico Nacional. 9-22, 2014.
- Miguel Jasso-Hernández, David Pinto, Darnes Vilariño. Análisis de sentimientos en Twitter: impacto de las características morfológicas. *Research in Computing Science*. Vol 72. ISSN: 1870-4069. Instituto Politécnico Nacional. 37-46, 2014.
- Mireya Tovar, David Pinto, Azucena Montes, Gabriel González, Darnes Vilariño, Beatriz Beltrán. Validación de conceptos ontológicos usando métodos de agrupamiento. *Research in Computing Science*. Vol 73. ISSN: 1870-4069. Instituto Politécnico Nacional. 9-18, 2014.
- Orlando Ramos, David Pinto, Belem Priego, Iván Olmos. Análisis empírico de la dispersión del español mexicano. *Research in Computing Science*. Vol 74. ISSN: 1870-4069. Instituto Politécnico Nacional. 9-20, 2014.
- Nahun Loya, Iván Olmos, David Pinto, Jesús González. Representación basada en grafos para la identificación de autoría en textos para el idioma español. *Research in Computing Science*. Vol 74. ISSN: 1870-4069. Instituto Politécnico Nacional. 119-128, 2013.
- Yuridiana Alemán, Darnes Vilariño, David Pinto. Clasificación de acosadores en línea utilizando Bootstrapping. *Research in Computing Science*. Vol 74. ISSN: 1870-4069. Instituto Politécnico Nacional. 139-150, 2013.
- María Beatriz Bernábe Loranca, David Pinto Avendaño, Jorge A. Ruiz-Vanoye, José Espinosa Rosales, Elías Olivares Benitez: A multi-objective proposal for the aggregation of economically inactive population, *International Journal of Combinatorial Optimization Problems and Informatics* 3(1): 70-79, 2012.
- Esteban Castillo, Darnes Vilariño, David Pinto, Maya Carrillo, Iván Olmos. Evaluación de características para el proceso de detección de autoría. *Research in Computing Science*. Vol 55. ISSN: 1870-4069. Instituto Politécnico Nacional. 2012.
- David Pinto, Beatriz Beltrán, Sofía Paniagua, Mireya Tovar, Darnes Vilariño: A Grammatical Inference System based on Genetic Algorithms, *Research in Computing Science Journal* 45(1): 355-364, 2010
- David Pinto, Mireya Tovar, Sofía Paniagua, Beatriz Beltrán, Darnes Vilariño: Some Experiments on Grammatical Inference of English, Spanish and Pseudo-English, *Research in Computing Science Journal* 45(1): 349-354, 2010
- David Pinto, Mireya Tovar, Beatriz Beltrán, Darnes Vilariño, Héctor Furlog: Feature Selection on the Web People Search Task, *Research in Computing Science Journal* 44(1): 77-85, 2009
- Franco Rojas López, Héctor Jiménez-Salazar, David Pinto, Aurelio López López: Dimensionality Reduction for Information Retrieval. *Research in Computing Science Journal* 20(1): 107-112, 2006
- Mireya Tovar, Maya Carrillo, David Pinto, Héctor Jiménez-Salazar: Combining Keyword Identification Techniques. *Research in Computing Science Journal* 14(1): 157-162, 2005
- Héctor Jiménez-Salazar, Mauricio Castro, David Pinto: Unsupervised Term Selection using Entropy. *Research in Computing Science Journal* 14(1): 163-172, 2005

(Otras)

- David Pinto: On Clustering and Evaluation of Narrow Domain Short-Text Corpora. *Procesamiento del Lenguaje Natural* 42(1): 129-130, 2009
- David Pinto, Alfons Juan, Paolo Rosso: A Comparative Study of Clustering Algorithms on Narrow-Domain Abstracts. *Procesamiento del Lenguaje Natural* 37(1): 43-49, 2006
- Héctor Jiménez-Salazar, David Pinto, Paolo Rosso: Uso del Punto de Transición en la Selección de Términos Índice para Agrupamiento de Textos Cortos, *Procesamiento del Lenguaje Natural* 35(1): 114-118, 2005

Lecture notes:



- Mireya Tovar, David Pinto, Azucena Montes, Gabriel González, Darnes Vilariño Ayala, Beatriz Beltrán: Use of Lexico-Syntactic Patterns for the Evaluation of Taxonomic Relations. MCPR 2014: 331-340, 2014
- Priego Sánchez Belém, David Pinto, Salah Mejri: Evaluating Polarity for Verbal Phraseological Units. MICAI (1) 2014: 191-200, 2014
- Helena Gómez-Adorno, David Pinto, Darnes Vilariño Ayala: A Question Answering System for Reading Comprehension Tests. MCPR 2013: 354-363, 2013
- Mireya Tovar, David Pinto, Azucena Montes, Darnes Vilariño Ayala: Determining the Degree of Semantic Similarity Using Prototype Vectors. MCPR 2013: 364-373, 2013
- David Pinto, Darnes Vilariño, Yuridiana Alemán, Helena Gómez, Nahun Loya and Héctor Jiménez-Salazar. The Soundex Phonetic Algorithm Revisited for SMS Text Representation. LNCS 7499, pp.47–55. 2012. Springer-Verlag, Berlín.
- Darnes Vilariño, David Pinto, Beatriz Beltrán, Saúl León, Esteban Castillo, and Mireya Tovar. A Machine-Translation Method for Normalization of SMS. MCPR 2012, LNCS 7329, pp. 293–302, 2012.
- Darnes Vilariño Ayala, David Pinto, Saúl León Silverio, Esteban Castillo and Mireya Tovar Vidal. BUAP: A Recursive Approach to the Data-Centric track of INEX 2011. LNCS 7424, pp. 161–166, 2012.
- Darnes Vilariño Ayala, David Pinto, Carlos Balderas, Mireya Tovar, Saul León: BUAP: A First Approach to the Data-Centric Track of INEX 2010. LNCS 6932, Springer-Verlag: 219-226, 2011. DOI: 10.1007/978-3-642-23577-1_19
- María Josefa Somodevilla García, Beatriz Beltrán, David Pinto, Darnes Vilariño Ayala, José Cruz Aaron: The BUAP Participation at the Web Service Discovery Track of INEX 2010. LNCS 6932, Springer-Verlag: 347-350, 2011. DOI: 10.1007/978-3-642-23577-1_33
- Mireya Tovar, Adrián Cruz, Blanca Vázquez, David Pinto, Darnes Vilariño Ayala, Azucena Montes: An Iterative Clustering Method for the XML-Mining Task of the INEX 2010, LNCS 6932, Springer-Verlag: 377-382, 2011, DOI: 10.1007/978-3-642-23577-1_36
- Darnes Vilariño, David Pinto, Carlos Balderas, Mireya Tovar, Beatriz Beltrán, and Sofia Paniagua: Use of Elliptic Curves in Term Discrimination. Advances in Pattern Recognition (MCPR 2011), LNCS 6718, Springer-Verlag: 341–349, 2011
- David Pinto, Darnes Vilariño, Carlos Balderas, Mireya Tovar, Beatriz Beltran: A Naïve Bayes Approach to Cross-Lingual Word Sense Disambiguation and Lexical Substitution. Advances in Pattern Recognition (MCPR 2010), LNCS 6256, Springer-Verlag: 352–361, 2010
- Fernando Perez-Tellez, David Pinto, John Cardiff, Paolo Rosso: Clustering Weblogs on the Basis of a Topic Detection Method. Advances in Pattern Recognition (MCPR 2010), LNCS 6256, Springer-Verlag: 342–351, 2010
- Darnes Vilariño, David Pinto, Mireya Tovar, Carlos Balderas, Beatriz Beltran: A Probabilistic Model Based on n-Grams for Bilingual Word Sense Disambiguation. Advances in Artificial Intelligence (MICAI 2010), LNCS 6437, Springer-Verlag: 82–91, 2010
- Antonio Juárez-González, Manuel Montes-y-Gómez, Luis Villaseñor Pineda, David Pinto Avendaño, Manuel Alberto Pérez-Coutiño: Selecting the N-Top Retrieval Result Lists for an Effective Data Fusion. CICLing 2010: 580-589, 2010
- Gabriela Ramírez-de-la-Rosa, Manuel Montes-y-Gómez, Luis Villaseñor Pineda, David Pinto Avendaño, Tamar Solorio: Using Information from the Target Language to Improve Crosslingual Text Classification. IceTAL 2010: 305-313, 2010
- David Pinto, Mireya Tovar, Darnes Vilariño, Beatriz Beltran, Héctor Jiménez-Salazar, Basilia Campos: BUAP: Performance of K-Star at the INEX'09 Clustering Task. INEX 2009. LNCS 6203, Springer-Verlag: 434-440, 2010
- David Pinto, Paolo Rosso, Héctor Jiménez-Salazar : On the Assessment of Text Corpora. NLDB 2009. LNCS 5723, Springer-Verlag: 281-290, 2010

- Fernando Perez-Tellez, David Pinto, John Cardiff, Paolo Rosso: Characterizing Weblog Corpora. NLDB 2009. LNCS 5723, Springer-Verlag: 299-300, 2010
- Rafael Guzmán-Cabrera, Paolo Rosso, Manuel Montes-y-Gómez, Luis Villaseñor Pineda, David Pinto Avendaño: Semi-supervised Word Sense Disambiguation Using the Web as Corpus. CICLing 2009: 256-265, 2009
- Fernando Perez-Tellez, David Pinto, John Cardiff, Paolo Rosso: Improving the Clustering of Blogosphere with a Self-term Enriching Technique. TSD 2009: 40-47, 2009
- Diego Ingaramo, David Pinto, Paolo Rosso, Marcelo Errecalde: Evaluation of Internal Validity Measures in Short-Text Corpora. CICLing 2008. Lecture Notes in Computer Science 4919, Springer-Verlag: 555-567, 2008
- David Pinto, Paolo Rosso: On the Relative Hardness of Clustering Corpora. TSD 2007. Lecture Notes in Artificial Intelligence 4629, Springer-Verlag: 155-161, 2007
- David Pinto, Alfons Juan, Paolo Rosso: Using Query-Relevant Documents Pairs for Cross-Lingual Information Retrieval. TSD 2007. Lecture Notes in Artificial Intelligence 4629, Springer-Verlag: 630-637, 2007
- David Pinto, José-Miguel Benedí, Paolo Rosso: Clustering Narrow-Domain Short Texts by Using the Kullback-Leibler Distance. CICLing 2007. Lecture Notes in Computer Science 4394, Springer-Verlag: 611-622, 2007
- Eugene Levner, David Pinto, Paolo Rosso, David Alcaide, R. R. K. Sharma: Fuzzifying Clustering Algorithms: The Case Study of MajorClust. MICAI 2007: 821-830, 2007
- Franco Rojas López, Héctor Jiménez-Salazar, David Pinto: A Competitive Term Selection Method for Information Retrieval. CICLing 2007. Lecture Notes in Computer Science 4394, Springer-Verlag: 468-475, 2007
- David Pinto, Héctor Jiménez-Salazar, Paolo Rosso: Clustering Abstracts of Scientific Texts Using the Transition Point Technique. CICLing 2006. Lecture Notes in Computer Science 3878, Springer-Verlag: 536-546, 2006
- David Pinto, Paolo Rosso, Ernesto Jiménez: A Penalisation Based-Ranked Approach for the Mixed Monolingual Task of WebCLEF 2006. CLEF 2006. Lecture Notes in Computer Science 4730, Springer-Verlag: 826-829, 2006
- Franco Rojas, Héctor Jiménez-Salazar, David Pinto: Vocabulary Reduction and Text Enrichment at WebCLEF. CLEF 2006. Lecture Notes in Computer Science 4730, Springer-Verlag: 830-834, 2006
- David Pinto, Héctor Jiménez-Salazar, Paolo Rosso, Emilio Sanchis: BUAP-UPV TPIRS: A System for Document Indexing Reduction at WebCLEF. CLEF 2005: 873-879, 2005

Memorias:

- Esteban Castillo, Ofelia Cervantes, Darnes Vilariño Ayala, David Pinto, Saul León: Unsupervised Method for the Authorship Identification Task. CLEF (Working Notes) 2014: 1035-1041, 2014
- Helena Gómez-Adorno, Grigori Sidorov, David Pinto, Alexander F. Gelbukh: Graph Based Approach for the Question Answering Task Based on Entrance Exams. CLEF (Working Notes) 2014: 1395-1403, 2014
- Yuridiana Alemán, Darnes Vilariño Ayala, David Pinto: Searching Sexual Predators in Social Network Dialogues. AMW 2013
- Helena Gómez-Adorno, David Pinto, Darnes Vilariño Ayala: Semantic Answer Validation in Question Answering Systems for Reading Comprehension Tests. AMW 2013
- Yuridiana Alemán, Nahun Loya, Darnes Vilariño Ayala, David Pinto: Two Methodologies Applied to the Author Profiling Task. CLEF (Working Notes) 2013
- Darnes Vilariño Ayala, David Pinto, Helena Gómez-Adorno, Saul León, Esteban Castillo: Lexical-Syntactic and Graph-Based Features for Authorship Verification Notebook for PAN at CLEF 2013. CLEF (Working Notes) 2013
- Helena Gómez-Adorno, David Pinto, Darnes Vilariño Ayala: Two Approaches for QA4MRE: Information Retrieval and Graph-based Knowledge Representation. CLEF (Working Notes) 2013

- Vivek Kumar Singh, Rajesh Piriyani, Ashraf Uddin, David Pinto: A Content-Based eResource Recommender System to Augment eBook-Based Learning. MIWAI 2013: 257-268, 2013
- David Pinto, Darnes Vilariño, Yuridiana Alemán, Helena Gómez, Nahún Loya. The Soundex Phonetic Algorithm revisited for SMS- based Information Retrieval. Pag 97-108. CERI 2012.
- David Pinto, Darnes Vilariño, Saúl León, Esteban Castillo, Mireya Tovar. Recuperación de Preguntas Frecuentes a través de SMS. XVI Congreso Español sobre Tecnologías y Lógica Fuzzy. 456-461. ESTYLF2012, Valladolid, España, 1-3 de febrero de 2012.
- Esteban Castillo, Darnes Vilariño Ayala, David Pinto, Ivan Olmos, Jesus A. Gonzalez, Maya Carrillo: Graph-based and Lexical-Syntactic Approaches for the Authorship Attribution Task. CLEF (Online Working Notes/Labs/Workshop) 2012
- Darnes Vilariño Ayala, Esteban Castillo, David Pinto, Ivan Olmos, Saul León: Information Retrieval and Classification based Approaches for the Sexual Predator Identification. CLEF (Online Working Notes/Labs/Workshop) 2012
- Mireya Tovar, J. Alejandro Reyes, Azucena Montes, Darnes Vilariño, David Pinto, Saúl León, BUAP: A First Approximation to Relational Similarity Measuring. First Joint Conference on Lexical and Computational Semantics (*SEM), 502–505, Montréal, Canada, June 7-8, (2012).
- Maya Carrillo; Darnes Vilariño; David Pinto; Mireya Tovar; Saul León; Esteban Castillo: FCC: Three Approaches for Semantic Textual Similarity. First Joint Conference on Lexical and Computational Semantics (*SEM), 631–634, Montréal, Canada, June 7-8, (2012)
- Darnes Vilariño; David Pinto; Mireya Tovar; Saul León; Esteban Castillo. BUAP: Lexical and Semantic Similarity for Cross-lingual Textual Entailment. First Joint Conference on Lexical and Computational Semantics (*SEM), 706–709, Montréal, Canada, June 7-8, (2012).
- María Beatriz Bernábe Loranca, Rogelio Gonzalez Velazquez, Elías Olivares Benitez, David Pinto Avendaño, Javier Ramírez Rodríguez, José Luis Martínez Flores: Partitioning with Variable Neighborhood Search: A bioinspired approach. NaBIC 2012: 150-155
- Darnes Vilariño Ayala, David Pinto, Saúl León Silverio, Esteban Castillo and Mireya Tovar Vidal: BUAP: A Recursive Approach to the Data-Centric track of INEX 2011. In pre-proceedings of the INEX 2011 Workshop. Shlomo Geva, Jaap Kamps, Ralf Schenkel, Andrew Trotman Eds., December 12–14, 2011, Hofgut Imsbach, Germany, <https://inex.mmci.uni-saarland.de/>, pp. 161-166, 2012.
- Darnes Vilariño Ayala, Esteban Castillo, David Pinto, Saul León, Mireya Tovar: Baseline Approaches for the Authorship Identification Task - Notebook for PAN at CLEF 2011. CLEF (Notebook Papers/Labs/Workshop) 2011
- María J. Somodevilla, Beatriz Beltrán, David Pinto, Darnes Vilariño, José Aaron. A first approach to web service discovery. In pre-proceedings of the INEX 2010 Workshop. Shlomo Geva, Jaap Kamps, Ralf Schenkel, Andrew Trotman Eds., December 13–15, 2010, Huize Bergen, Vught, the Netherlands, <http://www.inex.otago.ac.nz/>, pp. 286-289, 2010.
- Mireya Tovar Vidal, Adrian Cruz, Blanca Vázquez, David Pinto, Darnes Vilariño Ayala. An Iterative Clustering Method for the XML-Mining task of the INEX 2010. In pre-proceedings of the INEX 2010 Workshop. Shlomo Geva, Jaap Kamps, Ralf Schenkel, Andrew Trotman Eds., December 13–15, 2010, Huize Bergen, Vught, the Netherlands, <http://www.inex.otago.ac.nz/>, pp. 286-289, 2010.
- Darnes Vilariño Ayala, David Pinto, Carlos Balderas, Mireya Tovar. BUAP: A First Approach to the Data-Centric track of INEX 2010. In pre-proceedings of the INEX 2010 Workshop. Shlomo Geva, Jaap Kamps, Ralf Schenkel, Andrew Trotman Eds., December 13–15, 2010, Huize Bergen, Vught, the Netherlands, <http://www.inex.otago.ac.nz/>, pp. 159-168, 2010.

- Roberto Ortíz, David Pinto, Mireya Tovar, Héctor Jiménez-Salazar: BUAP: An Unsupervised Approach to Automatic Keyphrase Extraction from Scientific Articles. In proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010, pages 174–177, Uppsala, Sweden, 15-16 July 2010. Association for Computational Linguistics, ACL Anthology (<http://www.aclweb.org/anthology/S/S10/S10-1037.pdf>).
- Darnes Vilariño, Carlos Balderas, Miguel Rodríguez, David Pinto, Saúl León: FCC: Modeling Probabilities with GIZA++ for Task #2 and #3 of SemEval-2. In proceedings of the 5th International Workshop on Semantic Evaluation, ACL 2010, pages 112–116, Uppsala, Sweden, 15-16 July 2010. Association for Computational Linguistics, ACL Anthology (<http://www.aclweb.org/anthology/S/S10/S10-1023.pdf>).
- Esaú Villatoro-Tello, Luis Villaseñor-Pineda, Manuel Montes-y-Gómez, David Pinto-Avendaño: Multi-Document Summarization Based on Locally Relevant Sentences, 8th Mexican International Conference on Artificial Intelligence (MICAI'09), 87-91, 2009
- Fernando Perez Tellez, David Pinto, John Cardiff, Paolo Rosso: Defining and Evaluating Blog Characteristics, 8th Mexican International Conference on Artificial Intelligence (MICAI'09), 97-102, 2009
- David Pinto, Mireya Tovar, Beatriz Beltrán, Darnes Vilariño and Héctor Furlog: The Use of Document Fingerprinting in the Web People Search Task. SemWeb09. 534(1): 37-43 (2009)
- Jaime Lara, María de la Concepción Pérez de Celis and David Pinto: Dynamic Concept-Based Taxonomy used for image recovery based on their textual description . SemWeb09. 534(1): 28-36 (2009)
- David Pinto, Mireya Tovar, Darnes Vilariño, Beatriz Beltrán, Josefa Somodevilla: BUAP_1: A Naïve Approach to the Entity Linking Task. TAC 2009.
- Fernando Perez-Tellez, David Pinto, John Cardiff, Paolo Rosso: A Methodology to Cluster Informal Language Register Data. IICAI 2009: 1391-1401
- Alberto Barrón-Cedeño, Paolo Rosso, David Pinto, Alfons Juan: On Cross-lingual Plagiarism Analysis using a Statistical Model. PAN 2008, 9-13, 2008
- David Pinto, Paolo Rosso, Yassine Benajiba, Anas Ahachad, Héctor Jiménez-Salazar: Word Sense Induction in the Arabic Language: A Self-Term Expansion Based Approach. The Egyptian Society of Language Engineering (ESOLE): 235-245, 2007.
- David Pinto, Paolo Rosso, Héctor Jiménez-Salazar: UPV-SI: Word Sense Induction using Self-Term Expansion. 4th. Workshop on Semantic Evaluations - SemEval 2007. Association for Computational Linguistics, ACL Anthology: 430-433, 2007
- David Pinto, Paolo Rosso: KnCr: A Short-Text Narrow-Domain Sub-Corpus of Medline. TLH 2006. Advances in Computer Science: 266-269, 2006
- Sofia Paniagua Rivera, Héctor Jiménez-Salazar, David Pinto: An Unsupervised Method for Senses Clustering. IICAI 2005: 840-848, 2005

Tesis dirigidas concluidas (desde el año 2003):

Nivel maestría:

- Nombre: Un sistema de búsqueda semántica de información para su uso en el dominio de recuperación mejorada en yacimientos petroleros
- Fecha de examen: Diciembre/2013
- Nombre: Una metodología para el desarrollo de sistemas de búsqueda de respuestas orientados a pruebas de lectura comprensiva
Fecha de examen: Octubre/2013
- Nombre: Desarrollo de una metodología para la representación de texto mediante grafos etiquetados
Fecha de examen: Octubre/2013
- Nombre: Metodología para descubrir depredadores sexuales a partir del estudio de textos cortos
- Fecha de examen: Noviembre/2013
- Nombre: Una metodología para mitigar el fenómeno de plagio en el ámbito educativo
Fecha de examen: 03/03/2011

- Nombre: Herramienta de apoyo para el programa de formación docente
Fecha de examen: 19/11/2010
- Nombre: Aplicación de Métodos Basados en Corpus para la Tarea de Búsqueda de Personas en el Web
Fecha de examen: 11/08/2010
- Nombre: Métodos para la Generación de Extractos usando el Párrafo Virtual
Fecha de examen: 20/02/2006
- Nombre: Un nuevo Modelo de Ponderación para Sistemas de Recuperación de Información
Fecha de examen: 20/02/2006
- Nombre: Una Herramienta para el Preprocesamiento de Corpora
Fecha de examen: 20/02/2006
- Nombre: Una Metodología para la Construcción Automática de Thesauri Enriquecidos
Fecha de examen: 21/01/2005
- Nombre: Infraestructura Genérica de Flujo de Objetos para Aplicaciones Distribuidas bajo ProActive
Fecha de examen: 28/05/2004
- Nombre: Infraestructura Genérica para Agentes Móviles para Aplicaciones Distribuidas bajo ProActive
Fecha de examen: 28/05/2004
- Nombre: Obtención de Extractos de Textos con Base en un Corpus
Fecha de examen: 26/05/2004
- Nombre: Política de Fragmentación Horizontal Dinámica con Replicación Parcial de una Base de Datos
Fecha de examen: 20/01/2003

Nivel licenciatura:

- Nombre: Análisis empírico de la dispersión del español mexicano
Fecha de examen: Mayo/2014
- Nombre: Sistema de predicción de edad y género a partir de mensajes cortos
Fecha de examen: Agosto/2013
- Nombre: Reconocimiento automático de tópicos en documentos textuales
Fecha de examen: 23/08/2012
- Nombre: Plataforma open source para la integración de componentes en redes sociales
Fecha de examen: 27/05/2011
- Nombre: Detección Automática de Temas Importantes.
Fecha de examen: 06/10/2010
- Nombre: Corpus para un Diccionario de Ciencias de la Computación. Aplicación de Pruebas Piloto.
Fecha de examen: 11/02/2009
- Nombre: Recursos Lingüísticos de la Recuperación de Información en la FCC.
Fecha de examen: 06/11/2009
- Nombre: Sistema para Seguimiento de Órdenes de Servicio basado en Web.
Fecha de examen: 13/11/2009
- Nombre: Diseño y Construcción de un Sistema de Control Escolar para la Preparatoria Gral. Lázaro Cardenas del Río
Fecha de examen: 14/06/2004
- Nombre: Algoritmo de Expansión de Consultas Usando un Tesoro
Fecha de examen: 16/04/2004
- Nombre: Desarrollo e Implantación de una Heurística de Cerradura Transitiva para el Problema MaxSAT
Fecha de examen: 09/04/2003
- Nombre: Análisis y Diseño de un Sistema de Consultas Médicas con UML
Fecha de examen: 06/03/2003

Tesis dirigidas (en proceso):

Nivel doctorado:

- Nombre: Evaluación automática de ontologías de dominio
- Nombre: Identificación de expresiones verbales fijas para el español mexicano
- Nombre: Extracción de características basada en recorrido de caminos en grafos de texto

Nivel maestría:

- Nombre: Un sistema de visualización de ontologías para el dominio de extracción de hidrocarburos
- Nombre: Identificación de perfiles de usuario

Nivel licenciatura:

- Nombre: Aplicación de técnicas de inferencia gramatical para generación de expresiones
- Nombre: Implantación de un sistema manejador de base de datos para la gestión de archivos históricos
- Nombre: Recuperación de Información usando consultas tipo SMS

Participación en proyectos (desde el año 2003):

Título del proyecto: Incorporación de técnicas de tratamiento automático de información en humanoides

Entidad financiadora: CONACyT, Fondo de Infraestructura

Fecha de inicio: Marzo/2014

Fecha de término: Diciembre/2014

Investigador principal: Dr. David Eduardo Pinto Avendaño

Nivel de participación: Responsable

Título del proyecto: Machine learning techniques for context recommender systems

Entidad financiadora: CONACyT – Cooperación bilateral MEXICO-INDIA

Fecha de inicio: Enero/2013

Fecha de término: Diciembre/2016

Investigador principal: Dr. David Eduardo Pinto Avendaño

Nivel de participación: Responsable

Título del proyecto: Consultoría para la extracción de información en el dominio médico

Entidad financiadora: HDS - HEALTH DIGITAL SYSTEMS S.A.P.I. DE C.V

Fecha de inicio: 01/10/2012

Fecha de término: 15/03/2013

Investigador principal: Dr. David Eduardo Pinto Avendaño

Nivel de participación: Responsable

Título del proyecto: Aplicación de Técnicas de PLN en Procesos de Extracción Mejorada de Hidrocarburos

Entidad financiadora: Contrato de Servicios con el Instituto Mexicano del Petróleo

Fecha de inicio: 02/05/2011

Fecha de término: 14/03/2013

Investigador principal: Dr. David Eduardo Pinto Avendaño (BUAP)

Nivel de participación: Responsable

Título del proyecto: Evaluación Automática de Ontologías de Dominio Restringido (continuación)

Entidad financiadora: BUAP (VIEP)

Fecha de inicio: 01/03/2012

Fecha de término: 28/02/2013

Investigador principal: Dr. David Eduardo Pinto Avendaño (BUAP)

Nivel de participación: Responsable

Título del proyecto: Evaluación Automática de Ontologías de Dominio Restringido
Entidad financiadora: BUAP (VIEP)
Fecha de inicio: 01/03/2011
Fecha de término: 28/02/2012
Investigador principal: Dr. David Eduardo Pinto Avendaño (BUAP)
Nivel de participación: Responsable

Título del proyecto: Redes Sociales y la Web Semántica (Parte II)
Entidad financiadora: PROMEP-SEP
Fecha de inicio: 01/03/2011
Fecha de término: 28/02/2012
Investigador principal: Dr. David Eduardo Pinto Avendaño (BUAP)
Nivel de participación: Responsable

Título del proyecto: Estudio de Técnica Alternativas y de Alta Escalabilidad para Sistemas de Recuperación de Información
Entidad financiadora: CONACYT
Fecha de inicio: 01/01/2010
Fecha de término: 14/03/2014
Investigador principal: Dr. David Eduardo Pinto Avendaño (BUAP)
Nivel de participación: Responsable

Título del proyecto: Redes Sociales y la Web Semántica
Entidad financiadora: PROMEP-SEP
Fecha de inicio: 01/08/2009
Fecha de término: 31/07/2010
Investigador principal: Dr. David Eduardo Pinto Avendaño (BUAP)
Nivel de participación: Responsable

Título del proyecto: Extracción de relaciones léxicas para dominios restringidos a partir de contextos definatorios en español
Entidad financiadora: CONACYT
Fecha de inicio: 01/08/2008
Fecha de término: 31/07/2011
Investigador principal: Dr. Gerardo Sierra (UNAM)
Nivel de participación: Colaborador

Título del proyecto: Sistema de Búsqueda de Respuestas Inteligente basado en Agentes (ARA-ESP)
Entidad financiadora: Agencia Española de Cooperación Internacional (AECI)
Fecha de inicio: 01/01/2008
Fecha de término: 01/01/2009
Investigador principal: Dr. Paolo Rosso (Universidad Politécnica de Valencia, España)
Nivel de participación: Colaborador

Título del proyecto: TEXT-MESS: Minería de Textos Inteligente, Interactiva y Multilingüe basada en Tecnología del Lenguaje Humano (subproject UPV: MiDEs)
Entidad financiadora: CICYT TIN2006-15265-C06
Fecha de inicio: 01/10/2006
Fecha de término: 01/10/2009
Investigador principal: Patricio Martínez (Universidad de Alicante, España); MiDEs: Ferrán Pla (Universidad Politécnica de Valencia, España)
Nivel de participación: Colaborador

Título del proyecto: Almacenamiento y Búsqueda de Documentos usando Conceptos
Entidad financiadora: VIEP
Fecha de inicio: 15/03/2003
Fecha de término: 17/12/2003
Investigador principal: MC. David Eduardo Pinto Avendaño
Nivel de participación: Responsable

Título del proyecto: Infraestructura Genérica para Aplicaciones Distribuidas
Entidad financiadora: Facultad de Ciencias de la Computación
Fecha de inicio: 08/02/2002
Fecha de término: 14/06/2004
Investigador principal: MC. David Eduardo Pinto Avendaño
Nivel de participación: Responsable

Áreas de expertise:

- ***Tratamiento de información estructurada*** (Sistemas de bases de datos relacionales tipo Oracle, MySQL, Progress, Informix, etc)
- ***Tratamiento de información semi-estructurada*** (SGML, XML, HTML, Ontologías, SPARQL, etc)
- ***Tratamiento de información no estructurada*** (recuperación de información/máquinas de búsqueda, extracción de información, normalización de textos, clustering, machine learning, machine translation, resúmenes automáticos, compresión de textos, encriptación de información, lenguaje natural, SMS, redes sociales, creación, diseño y llenado de ontologías, etc)