

Alejandro Molina Villegas
PhD Informática, Investigador SNI
Procesamiento de Lenguaje Natural, Minería de Datos

EXPERIENCIA PROFESIONAL

- *CONACYT – Centro de Investigación en Ciencias de Información Geoespacial* Nov 2017 – Actual
Investigador Catedrático Conacyt.

- *Lean4it - Quarksoft* Feb 2017 – Nov 2017
Data Scientist - Proyectos especiales.

- *Comisión Nacional para el Conocimiento y Uso de la Biodiversidad* Sept 2014 – Ene 2017
Asesor de computo estadístico - minería de textos, Ecoinformática CONABIO.

- *Centre d’Enseignement et de Recherche en Informatique – UAPV (Francia)* 2011 – 2014
Investigador del *Laboratoire Informatique d’Avignon*.

- *Institut Universitaire de Technologie d’Avignon (Francia)* 2010 – 2011
Adjunto de investigación y docencia.

- *Facultad de Ingeniería de la UNAM* 2009 – 2010
Profesor de asignatura, Ingeniería en Computación.

- *3CT Training Center and Technology Consulting* 2009 – 2010
Web development trainer.

- *Centro de Estudios Técnicos Industriales y de Servicios* 2005 – 2007
Profesor de bachillerato.

GRADOS ACADÉMICOS

- Doctorado, *Informatique*
Mención honorífica
Université d’Avignon et des Pays de Vaucluse (UAPV), Francia 2013

- Maestría, Ciencias de la computación
Universidad Nacional Autónoma de México (UNAM), México 2009

- Licenciatura, Ciencias de la computación
Universidad Autónoma Metropolitana (UAM), México 2006

PROYECTOS

- *Georreferenciación Inteligente de Entidades en Discurso Libre (2017-en curso)*. Centro de Investigación en Ciencias de Información Geoespacial. Responsable técnico. Generación de modelos de machine learning y reconocimiento de patrones para decodificar el discurso, escrito o hablado para extraer coordenadas a partir de la detección de entidades nombradas. Compilación de corpus lingüísticos y gazetteers, desarrollo de módulos de software y publicación de artículos.

- *Misoginia en pocas palabras: Identificación y análisis de violencia escrita contra las mujeres en textos cortos de twitter (2018-en curso)*. Centro de Investigación en Ciencias de Información Geoespacial. Responsable técnico. Generación de modelos de detección de misoginia en redes sociales con métodos de procesamiento de lenguaje natural y análisis estadístico avanzado de covariables autoritativas relacionadas con la violencia contra las mujeres.
- *Sistema de monitoreo y análisis para la toma de decisiones en materia de congestión vehicular y movilidad para la CDMX (2016-2017)*. Secretaría de Ciencia, Tecnología e Innovación (SECITI) de la Ciudad de México. Responsable técnico. Recopilación, gestión, visualización y análisis de cantidades masivas de datos de reportes de Waze alineados con mapas de las vialidades de la zona metropolitana de la CDMX. Desarrollo de software con modelos predictivos del comportamiento del tráfico y publicaciones científicas.
- *Extracción automática de relaciones semánticas de especies Mexicanas (2014-2016)*. Comisión Nacional para el Conocimiento y Uso de la Biodiversidad. Responsable técnico. Coordinación del equipo de minería de texto y del desarrollo de la Biblioteca para la minería de textos de biodiversidad en español. Desarrollo de soluciones de gestión de bibliografía y OCR. Detección de patrones lingüísticos mediante expresiones regulares. Gestión de recursos, dirección de tesis de licenciatura y servicio social.
- *Detección y medición automática de similitud textual (2012-2014)*. Universidad Nacional Autónoma de México (México), Universitat Pompeu Fabra (España), Laboratoire Informatique d'Avignon (LIA) (Francia). Colaborador. Desarrollo de modelos computacionales para cuantificar la similitud semántica entre textos. Desarrollo de método de generación automática de resúmenes de textos basados en aprendizaje de máquina (machine learning).
- *Image on the Web: Analysis of the image life cycle through the Web 2.0. (2012-2015)*. XEROX, AMI Software, Université de Lyon, Électricité de France, Laboratoire Informatique d'Avignon, Centre d'Études Politiques de l'Europe Latine (Francia). Colaborador. Desarrollo de herramientas de anotación lingüística para la gestión de datos de clasificadores automáticos bayesianos. Publicaciones científicas y colaboración en equipo internacional.
- *Extracción léxica de contextos definitorios en dominios especializados (2008-2011)*. Grupo de Ingeniería Lingüística, UNAM. Tesis de Maestría. Desarrollo de método de clustering semántico de definiciones basado en técnicas probabilísticas. Programación de módulo de software en ambiente Web. Publicaciones científicas.

PUBLICACIONES ARBITRADAS

- [2020] Edwin Aldana-Bobadilla, Alejandro Molina-Villegas, Ivan López-Arévalo, Shanel Reyes-Palacios, Victor Muñoz-Sanchez, Jean Arreola-Trapala: Adaptive Geoparsing Method for Toponym Recognition and Resolution in Unstructured Text. *Remote. Sens.* 12(18): 3041 (2020).
- [2020] Melesio Crespo-Sánchez, Edwin Aldana-Bobadilla, Iván López-Arévalo and Alejandro Molina-Villegas: An Information-based Model for Writing Style Analysis of Lyrics. *First Workshop on NLP for Music and Audio NLP4MusA, ACL Anthology* (October, 2020).
- [2020] Saksik, R., Molina-Villegas, A., Linhares, A. C., & Torres-Moreno, J. M. Automatic Discourse Segmentation: an evaluation in French. *arXiv preprint arXiv:2002.04095*. (2020).
- [2019] Alcántara, F., Molina, A., & Muñoz, V. (2019, November). A Vector Semantics Approach to the Geoparsing Disambiguation Task for Texts in Spanish. In *Proceedings of the 1st International Con* (Vol. 13, pp. 47-55).

- [2019] Reyes-Palacios, S., Aldana-Bobadilla, E., Lopez-Arevalo, I., & Molina-Villegas, A. (2019, November). Georeference Assignment of Locations based on Context. In Proceedings of the 1st International Con (Vol. 13, pp. 39-46).
- [2019] Molina-Villegas, A., Siordia, O. S., Aldana-Bobadilla, E., Aguilar, C. A., & Acosta, O. (2019). Extracción automática de referencias geospaciales en discurso libre usando técnicas de procesamiento de lenguaje natural y teoría de la accesibilidad. *Procesamiento del Lenguaje Natural*, 63, 143-146.
- [2019] López P., Molina-Villegas A., Siordia O.: Geographical aggregation of microblog posts for LDA topic modeling. *Journal of Intelligent and Fuzzy Systems* 36(5): 4901-4908 (2019)
- [2019] Hernández-Chan, G., Villegas, A. M., Colunga, M. C., Siordia, O. S., & Rodríguez-González, A. (2019). Medic-Us: Advanced Social Networking for Intelligent Medical Services and Diagnosis. In *Current Trends in Semantic Web Technologies: Theory and Practice* (pp. 213-232). Springer, Cham.
- [2018] López-Ramírez, P., Molina-Villegas, A., Sánchez-Siordia, O., Chirinos-Colunga, M., & Hernández-Chan, G. (2018). Regular Activity Patterns in Spatio-Temporal Events Databases: Multi-Scale Extraction of Geolocated Tweets. *Research in Computing Science*, 147, 137-150.
- [2017] Aldana E., Lopez I., Molina A.: A Novel Data Reduction Method Based on Information Theory and the Eclectic Genetic Algorithm. *Intelligent Data Analysis Journal (JCR)*, ISSN: 1571-4128, Vol. 21 No. 4. (2017)
- [2016] Pérez C., Cortés J., Ramírez A., Abascal R., Molina A.: Recuperación, procesamiento y clasificación de tuits para visualizar estructuras de interacción. *Research in Computing Science*, ISSN: 1870-4069, Vol. 124 No. 1.
- [2016] Aldana E., Molina A.: A Novel Multivariate Mapping Method for Analyzing High-Dimensional Numerical Datasets. *Advances in Data Mining. Applications and Theoretical Aspects*, ISSN: 978-3-319-41560-4, pp 311-319
- [2016] Estrada R., Molina A., Perez-Espinosa A., Reyes A., Quiroz J., Bravo E. : Zonification of Heavy Traffic in Mexico City. *Proceedings of the 2016 International Conference on Data Mining (DMIN'16)*. Data Science, Springer. ISSN 2364-1185.
- [2015] Abascal-Mena R., Cossu J-V., Molina A., Torres-Moreno J-M.: Anotación automática de datos acerca de la reputación de los políticos en redes sociales. *Journal Research in Computing Science*. ISSN: 1870-4069.
- [2015] Molina, A., Torres-Moreno J.M.: El Test de Turing para la evaluación de resumen automático. *Linguamática, Revista para o Processamento Automático das Línguas Ibéricas* - ISSN: 1647-0818, Vol. 7 No. 2.
- [2015] Molina A., Sierra R., Zenteno D., Barrios J.: A Text Mining Library for Biodiversity Literature. *International Journal of Computational Linguistics and Applications*, ISSN 0976-0962, Vol. 6 No. 2.
- [2014] Velcin J., Brun C., Dormagen J. Y., Kim Y. M., Roux C., Boyadjian J., Molina, A.: Investigating the Image of Entities in Social Media: Dataset Design and First Results. *Proceedings of the European Language Resources Association (ELRA) conference*.
- [2014] Cossu J-V., Abascal-Mena R., Molina, A., Torres-Moreno J-M., SanJuan E.: Bilingual and cross domain politics analysis. *Journal Research in Computing Science*. ISSN: 1870-4069. Vol. 85. pp. 9-19.

- [2014] Cabrera-Diego A., Huet S., Jabaian B., Molina A., Torres-Moreno J.M., El-Bèze M., Durette B.: Algorithmes de classification et d'optimisation : participation du LIA/ADOC at DEFT'14. Proceedings of the Tenth DEFT Workshop.
- [2013] Molina A., Torres-Moreno J.M., da Cunha I., SanJuan E., Sierra G.: Discursive Sentence Compression. Proceedings of the 14th International Conference on Intelligent Text Processing and Computational Linguistics (CICLing'13). Lecture Notes in Computer Science. Springer. ISSN 0302-9743.
- [2013] Molina A., Sanjuan E., Torres-Moreno J.M. A TURING TEST TO EVALUATE A COMPLEX SUMMARIZATION TASK. Proceedings of Conference and Labs of the Evaluation Forum (CLEF'13).
- [2013] Saksik R., Molina A., Carneiro A., Torres-Moreno J.M.: Segmentação discursiva automática: uma avaliação preliminar em francês. Proceedings of the The 9th Brazilian Symposium in Information and Human Language Technology (STIL'13).
- [2012] Molina A., Torres-Moreno J.M., da Cunha I., Sanjuan E., Sierra G.: Sentence compression in Spanish driven by discourse segmentation and language models. En: Cornell University ArXiv: 1212.3493, Computation and Language (cs.CL), Information Retrieval (cs.IR), Vol.1212. 2012.
- [2011] Molina A., Torres-Moreno J.M., da Cunha I., Sanjuan E., Sierra G., Velázquez-Morales P.: Discourse Segmentation for Sentence Compression. En: Advances in Artificial Intelligence LNCS (Lecture Notes in Computer Science), Vol.7094, pp. 316-327. ISSN: 0302-9743.
- [2011] Cabrera-Diego L.A., Molina, A., Sierra, G.: A Dynamic Indexing Summarizer at the QA@INEX 2011 track. En: Initiative for the Evaluation of XML Retrieval Working Notes Series (INEX Question Answering Track'11).Springer-Verlag New York Inc. Vol.7424, pp. 154-159. ISBN 978-90-814485-8-1.
- [2010] Sierra G.,Torres-Moreno J.M., Molina A.: Regroupement sémantique de définitions en espagnol. In Proceedings of the 11th Atelier d'Evaluation des méthodes d'Extraction de Connaissances dans les Données (EGC/EVALECD'10),pp. 41–50. Hammamet, Tunes. 2010.
- [2010] Linhares A., Molina A., Torres-Moreno J.M., Peinl P.: Usando Grafos e Algoritmo de Karp na Sumarização Automática de Documentos. En: Proceedings of the 42th Simpósio Brasileiro de Pesquisa Operacional (SBPO'10). Rio de Janeiro. 2010.
- [2010] Molina A., Sierra G., Torres-Moreno J.M.: La energía textual como medida de distancia en agrupamiento de definiciones. En: Proceedings of the 10th Journées Internationales d'Analyse statistique des Données Textuelles (JADT'10). Vol.3, pp. 215-226. Roma. 2010.ISBN: 978-88-7916-450-9
- [2010] Molina, A., da Cunha, I., Torres-Moreno, J.M., Velázquez-Morales P.:La compresión de frases: un recurso para la optimización de resumen automático de documentos. En: Journal Lingüística, Vol.2, num. 3, pp.13–27. 2010. ISSN: 1647-0818.
- [2010] da Cunha I., Molina A., Velázquez-Morales P., Torres-Moreno J.M.: Optimización de resumen automático mediante compresión de frases. En: Proceedings of the 28th Congreso Internacional de la Asociación Española de Lingüística Aplicada (AESLA'10). Vol.28, pp. 73-83. Vigo, España. 2010. ISBN: 978-84-8158-479-0.
- [2009] Sierra G., Alarcón R., Molina A. AldanaE.: Web Exploitation for Definition Extraction. IEEE Latin American Web Congress (LA-WEB'09).pp. 217-223. 2009. ISBN: 978-0-7695-3856-3.

FORMACIÓN DE RECURSOS HUMANOS

- Director: *Técnicas de Machine Learning para detección de discursos de odio*. Jonathan Zárate Cartas, Centro de Investigación en Ciencias de Información Geoespacial (Centrogeo) – Doctorado en Ciencias de Información Geoespacial. En curso.
- Director: *Acceso a la información y gobernanza digital*. Hermelando Cruz Pérez, Centro de Investigación en Ciencias de Información Geoespacial (Centrogeo) – Doctorado en Ciencias de Información Geoespacial. En curso.
- Director: *Aproximación al análisis de la violencia contra las mujeres mediante ciencias de datos*. Ana Laura Bernal Gómez, Centro de Investigación en Ciencias de Información Geoespacial (Centrogeo) – Maestría en Ciencias de Información Geoespacial. En curso.
- Co-director: *Técnicas de Deep Learning para Resumen Automático en Español*. Ana Beatriz Rodríguez Mendoza, Centro de Investigación en Matemáticas (CIMAT) – Maestría en Computo estadístico. En curso.
- Co-director: *Reconocimiento de Entidades Nombradas Georeferenciables con Word Embeddings*. Jean Michel Arreola Trapala, Centro de Investigación en Matemáticas (CIMAT) – Maestría en Computo estadístico. 2019.
- Co-director: *Métodos de Desambiguación para Geoparsing en Textos en Español*. Filomeno Alberto Alcántara Alonso, Centro de Investigación en Matemáticas (CIMAT)–Maestría en Computo estadístico). 2019.
- Co-director: *Aplicación de técnicas de minería de datos para la zonificación del tráfico de la ciudad de México*. Fernando Sandoval (Universidad Autónoma de la Ciudad de México (UACM)–Ingeniería en Software). En curso.
- Asesor: *Extracción terminológica automática y agrupamiento de términos semejantes mediante coocurrencias e información mutua*. Alumno: Armando Avila. Universidad Nacional Autónoma de México–Maestría en Ciencias de la Computación. 2016.
- Jurado: *Comunicación Política en Twitter: Análisis estructural de interacciones entre Sistema Político, Sistema de Medios y Ciudadanos mexicanos en tres eventos de Comunicación Política*. Alumnos: Jorge Cortés Rodríguez, Alejandra Martínez Kuri, Carlos Pérez Cerón, Anaíd Simón Ávila (Universidad Autónoma Metropolitana–Maestría en Diseño, información y comunicación). 2016.
- Responsable de Servicio social: *Minería de datos para la extracción de información de las especies de México*. Alumnos: Alejandra Yvette Ramos, Mayra Asli Perez (Universidad Nacional Autónoma de México; Universidad Autónoma Metropolitana–Licenciatura en Biología). 2016
- Responsable de Servicio social: *Detección y desambiguación de nombres, científicos y comunes, de especies mexicanas*. Alumnos: David Flores, Jesús Medina y Edgar Moreno (Universidad Autónoma Metropolitana–Licenciatura en Ciencias de la Computación). 2015.
- Co-director: *Segmentation Discursive Automatique*. Rémy Saksik (UAPV–bac+3 informatique). 2014.

DOCENCIA

- *Miembro del Núcleo Académico Básico*. Programa: Maestría en Geointeligencia Computacional. Centrogeo.
- *Seminaire de Projet Final*. Programa: Licence Informatique (UAPV–Licenciatura), 23 horas.
- *Algorithmique et Optimisation*. Programa: Licence Informatique (UAPV–Licenciatura), 21 horas.
- *Conception Internet et Bases de Données*. Programa: Licence Informatique (UAPV–Licenciatura), 60 horas.
- *Programmation de Systèmes d'Exploitation*. Programa: Licence Informatique (UAPV–Licenciatura), 48 horas.
- *Conception Internet*. Programa: Licence Informatique (UAPV–Licenciatura), 66 horas.
- *Web Dynamique*. Programa: Licence Informatique (UAPV–Licenciatura), 42 horas.
- *Bases de la Programmation*. Programa: Licence Informatique (UAPV–Licenciatura), 18 horas.
- *Modelisation Objet et UML*. Programa: Licence Informatique (UAPV–Licenciatura), 9 horas.
- *Environnement Informatique et Bureutique*. Programa: Statistique et Informatique Décisionnelle (Institut Universitaire de Technologie d'Avignon–Licenciatura), 33 horas.

- *Laboratorio de computación gráfica*. Programa: Ingeniería en Computación (Facultad de ingeniería de la UNAM–Licenciatura), 60 horas.
- *Programación en PHP*. Centro de Capacitación y Consultoría Tecnológica, más de 80 horas.
- *Bases de datos relacionales en MySQL*. Centro de Capacitación y Consultoría Tecnológica, más de 80 horas.
- *HTML con hojas de estilo CSS*. Centro de Capacitación y Consultoría Tecnológica, más de 80 horas.
- *Ciencias de la computación y Matemáticas*. Centro de Estudios Tecnicos Industriales y de Servicios–Bachillerato Técnico, más de 100 horas.

RECONOCIMIENTOS

- Miembro del Sistema Nacional de Investigadores SNI.
- Tercer lugar en el certamen de detección automática de textos agresivos MEX-AT3.
- Miembro fundador de la Asociación Mexicana de Procesamiento del Lenguaje Natural.
- Ex-Coordinador del Grupo de Análisis y Reconocimiento de Voz de la Red Temática en Tecnologías de Lenguaje.
- Beca doctoral en el extranjero. CONACYT–Embajada de Francia.
- Promedio más alto de la generación 2006 en examen CENEVAL – UAM-I.
- Representante del consejo 2005 de ciencias básicas e ingeniería. UAM-I .
- Tercer lugar de proyectos científicos bachillerato. CETIS-37.

IDIOMAS

- Inglés: TOEFL Score: 537 + Diploma de la Coordinación de Enseñanza de Lenguas Extranjeras (CELEX-UAM) por el curso de dominio de Inglés.
- Francés: DELF B2 session 2009 + seis 6 radicando en Francia.

HABILIDAD TÉCNICA

Python, TensorFlow, Scikit-Learn, R, HTML/XML, SQL, Bash, Regex, MongoDB, Git, Unix/Linux.