Enrique Noriega-Atala

Phone: +1 (520) 241-5945
Email: enrique.noriega@gmail.com
Website: https://enoriega.info

Education

Ph.D. Information Science, The University of Arizona (2020).
- Minor: Statistics

M.S. Computer Science, The University of Arizona (2014).

B.S. Information Technologies, Monterrey Institute of Technology and Higher Education (2010).

Work Experience

Oct 2022 - Currently Computational Sciences Researcher IV, Computer Science Dept., The University of Arizona (UA)

My primary duty is to conduct research in the field of Natural Language Processing (NLP) and collaborate with interdisciplinary teams from different academic and industry institutions with the goal of producing peer-reviewed publications and collaborate in the procurement of extramural funding. I also serve as a close collaborator in the Data Sciente Institute (DSI), where I perform as a domain matter expert in Large Language Models (LLM) and liaison between DSI and faculty and research staff across campus to support and consult in their research projects. During my tenure in this position, I participated as a co-Principal Investigator for the *SKEMA* research project, a multi-year effort part of DARPA's ASKEM program and participate in the curriculum design and teaching of AI related internal workshops.

Responsibilities:

- Consult as NLP/LLM domain matter expert in research engagements between DSI and various academic departments.
- Co-develop curriculum and teach workshops for DSI's DataLab. Examples include *Exploring the LLM Frontier* and *AI Makerspace*.
- Leading and collaborating on research projects about information retrieval, extraction, document modeling, program synthesis, etc.
- Writing and publishing research work on top-tier, peer-reviewed venues.
- Collaborating with planning and writing extramural research grants and other funding mechanisms.
- Participating in the organization of academic workshops.
- Mentoring graduate students.

Jan 2021 - Sept 2022 Postdoctoral Research Associate, Computer Science Dept., The University of Arizona (UA)

Conduct research focused on biomedical NLP and collaborate with interdisciplinary teams within the health and computing sciences academic units, under the umbrella of the *Collaborative for Global Adaptive Pandemic Solutions* initiative, aimed at creating models and interfaces to explore the vast biomedical scientific literature.

Achievements:

- Collaborated on the development of five research publications, accepted and under review to top-tier peer-reviewed venues.
- Lead a collaborative effort to develop a multi-disciplinary visual analytics tool for the visualization of information extraction in the biomedical domain.
- Contributed to open source NLP projects.
- Presented our research projects in multiple symposiums and conferences.
- Collaborated on the development of two grant proposals, for NSF and DARPA.
- Collaborated and coordinated the organization of two research workshops, one institutional and the other co-hosted in Coling '22.
- Mentored graduate students in the dept. of Computer Science.

Feb 2020 - Oct 2020 Senior NLP Scientist, CodaMetrix

Contributed to the research and development of Natural Language Processing algorithms and models for electronic health records (EHR).

Some of the specific projects to which I contributed include clinical information extraction tools, such as an anatomy phrase classifier or a laterality named-entity recognizer, tailored for radiology and pathology clinical notes.

The tasks involved include:

- Building a taxonomy of body regions, organs and tissues using medical ontologies such as UMLS, SNOMED and MeSH.
- Bootstrapping a dataset using proprietary EHR corpora.
- Implementing, training and testing state-of-the-art neural architectures with PyTorch and Tensor Flow.
- Developing APIs for integration with the company's product.

In addition to research and development, I participated as a mentor in the summer internship program and contributed to the creation of documentation and internal training curricula NLP and machine learning.

Jan 2013 - Jan 2020 Graduate Research Associate, UA

Worked within CLU lab and ML4AI on multiple Natural Language Processing (NLP) and Machine Learning projects applied to the health sciences.

The focus of my contributions has been on the areas of Information Extraction, Information Retrieval and Reinforcement Learning and has lead to multiple publications in top-tier peer-reviews venues.

During my graduate work, I've been a regular contributor to REACH, a machine reading system for biomedical publications.

Aug 2013 - Dec 2019 Teaching Assistant, UA

During my time as a graduate student, I collaborated as a teaching assistant for multiple graduate and undergraduate courses both in the Department of Computer Science and the School of Information. My duties consisted primarily of grading assignments, exams and projects as well as holding office hours for students. Also, occasionally, included designing assignments, directing laboratory sessions and guest lectures.
Courses:

- Computer Organization (2013)
- Data Structures (2014)
- Information Retrieval and Web Search (2015, 2016)
- Introduction to Machine Learning (2017, 2019)

2010 - July 2012 Sr Software Developer, FSC

Worked on the development of an ERP system and multiple ancillary products for the produce industry based on the .NET stack.

Lead the migration of multiple enterprise information systems into Microsoft's Azure cloud from their previous on-premises location. Designed and coordinated the migration of software assets and data, as well as worked on software adaptations to make the systems work properly on their new cloud-based home.

Participated on the design and architecture of multiple information systems based on the .NET stack.

2009 - 2010 Software Developer, Teknol, SA de CV

Worked on GIS systems for multitouch hardware for the mining industry.

2008 - 2009 Internship at Centro de Investigación y Desarrollo de Ingeniería Avanzada, AC

Web developer working with Python and Django-based projects.

2007 - 2008 Web Developer, Optima Commerce LLC

Contributed in the development of web-based international trade information system based on ASP.NET.

2007 Jr. Developer, Marketing Movil S.A de C.V

Contributed in the development of a Java-based service for sending and transmitting text messages (SMS) for a local marketing agency.

Honors and Awards

UA Graduate & Professional Student's Council Travel Grant (2018).

UA School of Information's Travel Award (2018).

International Conference in Data Mining Student Award (2018).

Galileo Circle Scholar, UA's College of Science (2014).

Instituto Educativo Sonora-Arizona scholarship (2013-2014).

CONACyT graduate studies scholarship (2012-2014).

CENEVAL's outstanding performance testimony on the EGEL test (2010).

Presentations and Publications

Journal Publications

Enrique Noriega-Atala, Paul D. Hein, Shraddha S. Thumsi, Zechy Wong, Xia Wang, Sean M. Hendryx, Clayton T. Morrison
Extracting Inter-sentence Relations for Associating Biological Context with Events in Biomedical Text. *IEEE/ACM transactions on computational biology and bioinformatics. 2019.*

Marco A. Valenzuela-Escárcega, Özgün Babur, Gus Hahn-Powell, Dane Bell, Thomas Hicks, **Enrique Noriega-Atala**, Xia Wang, Mihai Surdeanu, Emek Demir, Clayton T. Morrison.
Large-scale Automated Machine Reading Discovers New Cancer Driving Mechanisms. *Database: The Journal of Biological Databases and Curation. 2018.*

Conference Publications

Enrique Noriega-Atala, Robert Vacareanu, Salena Torres Ashton, Adarsh Pyarelal, Clayton T Morrison and Mihai Surdeanu. When and Where Did it Happen? An Encoder-Decoder Model to Identify Scenario Context.
Findings of The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP) 2024.

Robert Vacareanu, **Enrique Noriega-Atala**, Gus Hahn-Powell, Marco A Valenzuela-Escárcega and Mihai Surdeanu.
Active learning design choices for NER with transformers.
Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, (LREC-COLING) 2024.

Enrique Noriega-Atala, Md Rahat-Uz-Zaman, Ruchika Bhat, Mladen Jergović, Stephen G. Kobourov, Janko Nikolich-Žugich.
Visualizing Interaction Networks and Evidence in Biomedical Corpora.
Proceedings of the 16^{th} *IEEE Pacific Visualization Symposium*, (PacificVis) 2023.

Robert Vacareanu, George C Barbosa, **Enrique Noriega-Atala**, Gus Hahn-Powell, Rebecca Sharp, Marco A Valenzuela-Escárcega, Mihai Surdeanu.
A Human-machine Interface for Few-shot Rule Synthesis for Information Extraction.
Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, (NAACL-HLT) 2022.

Enrique Noriega-Atala, Marco A. Valenzuela-Escárcega, Clayton T. Morrison and Mihai Surdeanu. Learning what to read: Focused machine reading. *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (EMNLP), 2017.

Workshop Publications

Enrique Noriega-Atala, Robert Vacareanu, Gus Hahn-Powell and Marco A. Valenzuela-Escárcega.
Neural-Guided Program Synthesis of Information Extraction Rules Using Self-Supervision. *Proceedings of the Workshop on Pattern-based Approaches to NLP in the Age of Deep Learning at COLING, Gyeongju, Republic of Korea, 2022.*

Enrique Noriega-Atala, Mihai, Surdeanu, Clayton T. Morrison.
Learning Open Domain Multi-hop Search Using Reinforcement Learning. *In Proceedings of the Workshop on Structured and Unstructured Knowledge Integration, Seattle, Washington. Association for Computational Linguistics, 2022.*

Zhengzhong Liang, **Enrique Noriega-Atala**, Clayton Morrison, Mihai Surdeanu.
Low Resource Causal Event Detection from Biomedical Literature.
Proceedings of the 21st Workshop on Biomedical Language Processing at ACL, 2022.

Enrique Noriega-Atala, Peter M. Lovett, Clayton T. Morrison, Mihai Surdeanu.
Neural Architectures for Biomedical for Biological Inter-Sentence Relation Extraction. *Proceedings of the Scientific Document Understanding at AAAI, 2022*

Enrique Noriega-Atala, Zhengzhong Liang, John A. Bachman, Clayton T. Morrison, Mihai Surdeanu.
Understanding the Polarity of Events in the Biomedical Literature: Deep Learning vs. Linguistically-informed Methods. *Proceedings of the Workshop on Extracting Structured Knowledge from Scientific Publications at NAACL-HLT, 2019.*

Enrique Noriega-Atala, Paul D. Hein, Shraddha S. Thumsi, Zechy Wong, Xia Wang, Clayton T. Morrison.
Inter-sentence Relation Extraction for Associating Biological Context with Events in Biomedical Texts. *Proceedings of the Sixth Workshop on Data Mining in Biomedical Informatics and Healthcare at ICDM, 2018.*

Enrique Noriega-Atala, Marco A. Valenzuela-Escáarcega, Clayton T. Morrison and Mihai Surdeanu.
Focused Reading: Reinforcement Learning for What Documents to Read. *Proceedings of the Interactive Machine Learning and Semantic Information Retrieval Workshop at ICML, 2017.*

Marco A. Valenzuela-Escárcega, Özgün Babur, Gus Hahn-Powell, Dane Bell, Thomas Hicks, **Enrique Noriega-Atala**, Xia Wang, Mihai Surdeanu, Emek Demir, Clayton T. Morrison.
Large-scale Automated Reading with Reach Discovers New Cancer Driving Mechanisms. *Proceedings of the BioCreative VI Workshop* (BioCreative6 2017), 2017, pp. 200-202.

Conference Presentations

When and Where Did it Happen? An Encoder-Decoder Model to Identify Scenario Context. *Workshop on the Future of Event Detection at EMNLP 2024*, Miami, Florida (2024).

Visualizing Interaction Networks and Evidence in Biomedical Corpora. *The* 16^{th} *IEEE Pacific Visualization Symposium*, (PacificVis), Seoul, South Korea (2023).

Neural-Guided Program Synthesis of Information Extraction Rules Using Self-Supervision. *Workshop on Pattern-based Approaches to NLP in the Age of Deep Learning at COLING*, Gyeongju, Republic of Korea (2022).

Open Domain Multi-hop Search Using Reinforcement Learning. *Workshop on Structured and Unstructured Knowledge Integration at NAACL-HLT*, Seattle, Washington (2022).

Low Resource Causal Event Detection from Biomedical Literature. *21st Workshop on Biomedical Language Processing at ACL*, Dublin, Ireland (2022).

Neural Architectures for Biomedical for Biological Inter-Sentence Relation Extraction. *Scientific Document Understanding at AAAI*, Vancouver, Canada (2022).

Understanding the Polarity of Events in the Biomedical Literature: Deep Learning vs. Linguistically-informed Methods. *Workshop on extracting structured knowledge from scientific publications at NAACL-HLT*, Minneapolis, Minnesota (2019).

Inter-sentence Relation Extraction for Associating Biological Context with Events in Biomedical Text. *Sixth Workshop on Data Mining in Biomedical Informatics and Healthcare at ICDM*, Singapore (2018).

Grounding Gradable Adjectives through Crowd-sourcing *Language Resources and Evaluation Conference*, Miyazaki, Japan (2018).

Learning what to read: Focused machine reading. *NLP Applications track, EMNLP*, Copenhagen, Denmark (2017).

Other Presentations and Guest Lectures

Resilience Informatics Workshop: *AZX Chat Bot: An LLM-driven application implementation to support resilience against climate change*. The University of Arizona, November 20^{th}, 2023

Immune Monitoring Symposium: *Visualizing Interaction Networks in Evidence in Biomedical Corpora*. The University of Arizona, November 7^{th}, 2023.

PHIRE Side Chat: *Machine Leaning and NLP for Medical Research and Public Health*. The University of Arizona, May 25^{th}, 2023.

MCB Joint Seminar *Accelerating Research through Large Scale, Interpretable Machine Reading*. The University of Arizona, October 11^{th}, 2022.

Applied Data Science Speak & Greet Speaker Series *Accelerating Research through Large Scale, Interpretable Machine Reading*. The University of Arizona, September 21^{st}, 2022.

What is Machine Learning?. For the course *Knowledge-based Systems in the Organizations. Systems and Industrial Engineering program*. Monterrey Institute of Technology and Higher Education, Hermosillo, México, 2015.

Service

Organization Committee Member:

The Second Workshop in Pattern-based Approaches to NLP in the Age of Deep Learning (Pan-DL), EMNLP, 2023.

Pattern-based Approaches to NLP in the Age of Deep Learning (Pan-DL), COLING, 2022.

Program Committee Member:

The AAAI-22 Workshop on Scientific Document Understanding., AAAI, 2022

SUKI: Structured and Unstructured Knowledge Integration, NAACL-HLT, 2022.

TextGraphs: Graph-based Algorithms for Natural Language Processing, 14, 15 and 16.

Student volunteer at NAACL-HLT. Minneapolis, Minnesota (2019).

Grants

Co-PI: *SKEMA: Scientific Knowledge Extraction and Model Analysis*, DARPA.
Duration: July 2022 through December 2025 (award amount $9,345,747).

Key Personnel: *Resilience Informatics for Public Health*, Technology and Research Initiative Fund (TRIF), The University of Arizona.
Duration: July 2023 through June 2024 (award amount $95,000).

Patents

Mihai Surdeanu, Marco A. Valenzuela-Escarcega, Gustave Hahn-Powell, Dane Bell, Thomas Hicks, **Enrique Noriega**, Clayton Morrison. "Method for Extracting and Assessing Signaling Pathways from Biomedical Literature". Provisional US patent filed on 03/2017.

Skills

Programming Languages:

Python, Scala, Java, C#, C, C++, R, Matlab, JavaScript, VB.NET.

Scientific Tooling:

PyTorch, DyNET, Scikit-Learn, NumPy, SciPy, Matplotlib, CoreNLP, REACH.

Databases

Relational: SQL Server, Postgres. MySQL. NoSQL: Redis, Neo4J.

IDEs

IntelliJ, PyCharm, Visual Studio, Netbeans.

CI/CD

Travis CI, Maven, SBT, Makefiles.

Cloud

AWS: Compute. Azure: AppServices, VMs, Service Bus, Storage, Networking.

Systems administration

Unix-like environment proficiency. Bash scripting.