



Revista Electrónica de  
Tecnología, Educación y Ciencia  
ISSN: 2953-5654  
<http://retec.unsa.edu.ar>  
Universidad Nacional de Salta

**Trabajo final de grado**

**Análisis y Diseño de una herramienta software para  
Reconocimiento y Clasificación de entidades nombradas en  
grandes volúmenes de textos históricos**

**Facundo Darfe**

Facultad de Ciencias Exactas, Universidad Nacional de Salta  
[facudarfe@gmail.com](mailto:facudarfe@gmail.com)

**Revista Electrónica de Tecnología, Educación y Ciencia,  
Volumen 1, Número 1, pág. 82-84, jun, 2023. ISSN: 2953-5654**

Disponible en <http://retec.unsa.edu.ar/>

Trabajo final de grado

**Análisis y Diseño de una herramienta software para  
Reconocimiento y Clasificación de entidades nombradas en  
grandes volúmenes de textos históricos**

Facundo Darfe

Facultad de Ciencias Exactas – Universidad Nacional de Salta  
[facudarfe@gmail.com](mailto:facudarfe@gmail.com)

**Director:** Dr. Eduardo Xamena (UNSa)

**Co-Director:** Lic. Ismael Orozco

**Carrera:** Licenciatura en Análisis de Sistemas

**Año:** 2023

## Objetivo

El proyecto tuvo como objetivo el diseño y desarrollo de una herramienta de software para el reconocimiento y clasificación de entidades nombradas (NERC) en grandes volúmenes de textos históricos, con un enfoque específico en documentos relacionados con la vida del General Martín Miguel de Güemes.

Se utilizó la metodología de Investigación en Ciencias de la Computación (DSRM) para lograr un rendimiento competitivo en la tarea de NERC. La metodología plantea fases iterativas de diseño, desarrollo, demostración, evaluación y comunicación de resultados que se aplicaron de manera rigurosa al trabajo. El proyecto también incluye la creación de un conjunto de datos y la evaluación de diferentes herramientas y técnicas de aprendizaje automático para lograr los objetivos establecidos.

## Resumen del trabajo

La identificación de entidades nombradas es esencial para la extracción de información y la búsqueda de interacciones entre personajes, fechas, hechos y otros tipos de entidades en los documentos históricos. Esto se logra mediante el uso de herramientas de Procesamiento de Lenguaje Natural (NLP) y Reconocimiento y Clasificación de Entidades Nombradas (NERC).

El proyecto se centra en el problema del reconocimiento y clasificación de entidades nombradas en los documentos titulados como “Güemes Documentado” que relatan la vida del General Martín Miguel de Güemes. Se utilizaron una combinación de técnicas de aprendizaje automático, incluyendo redes neuronales convolucionales (CNN) y transformers, para identificar y clasificar entidades nombradas en el texto como pueden ser personas, fechas y lugares. El proyecto también implica la creación de un conjunto de datos y la evaluación de diferentes herramientas y técnicas de machine learning para lograr los objetivos establecidos.

A lo largo del trabajo se utilizó la metodología de Investigación en Ciencias de la Computación (DSRM), que es un proceso para diseñar soluciones de software para la investigación científica. La metodología consta de seis fases, con la primera fase siendo la identificación del problema y la motivación para encontrar una solución. La segunda fase implica definir los objetivos para la solución. Las fases restantes implican diseño iterativo, desarrollo, demostración, evaluación y comunicación de los resultados. La metodología enfatiza la importancia de la rigurosidad tanto en el desarrollo como en la evaluación de la solución, así como la comunicación efectiva de los resultados al público adecuado

## Referencias

1. Transformers en procesamiento del lenguaje natural - IIC, Mayo 2022.
2. Douglas Appelt, Jerry R Hobbs, John Bear, David Israel, Megumi Kameyama, Andrew Kehler, David Martin, Karen Myers, and Mabry Tyson. Sri international fastus systemmuc-6 test results and analysis. In Sixth Message Understanding Conference (MUC-6): Proceedings of a Conference Held in Columbia, Maryland, November 6-8, 1995, 1995.
3. José Canete, Gabriel Chaperon, Rodrigo Fuentes, Jou-Hui Ho, Hojin Kang, and Jorge Pérez. Spanish pre-trained bert model and evaluation data. Pml4dc at iclr, 2020:1–10, 2020.
4. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.
5. Anton A Emelyanov and Ekaterina Artemova. Multilingual named entity recognition using pretrained embeddings, attention mechanism and ncrf. arXiv preprint arXiv:1906.09978, 2019.
6. LUIS GUEMES. Güemes documentado. 1815.
7. Karen Yanet Haag. Reconocimiento de entidades nombradas en texto de dominio legal. B.S. thesis, 2019.
8. Ji Young Lee, Franck Dernoncourt, and Peter Szolovits. Transfer learning for named-entity recognition with neural networks. arXiv preprint arXiv:1705.06273, 2017.
9. Alireza Mansouri, Lilly Suriani Affendey, and Ali Mamat. Named entity recognition approaches. International Journal of Computer Science and Network Security, 8(2):339–344, 2008.
10. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. arXiv preprint arXiv:1706.03762, 2017.
11. R Hevner Von Alan, Salvatore T March, Jinsoo Park, and Sudha Ram. Design science in information systems research. MIS quarterly, 28(1):75–105, 2004.
12. Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, R´emi Louf, Morgan Funtowicz, et al. Transformers: State-of-the-art natural language processing. In Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations, pages 38–45, 2020.
13. Shi Yan. Understanding lstm and its diagrams. MLReview. com, 2016.