ELENA ZOTOVA

NLP Researcher

Donostia San Sebastián, Spain
linkedin.com/in/elena-zotova-03825183   github.com/ZotovaElena   orcid.org/0000-0002-8350-1331

EXPERIENCE

NLP Researcher
Vicomtech

March 2020 – Present   Donostia/San Sebastián

- Research in European and national projects, funded by government and industry
- Development of classification models, topic modeling, clustering, summarization, and entity recognition in Spanish, English, Basque and other European languages
- Information retrieval for question answering systems, search, and RAG
- LLMs fine-tuning (with quantization)
- PhD thesis *Multilingual Information Extraction in Clinical Texts Using Deep Learning Approaches*, focused on named entity recognition and entity linking in electronic health records and clinical terminologies.
- Scraping and processing of unstructured text
- Data annotation for machine learning systems
- Implementation of web services

Computational Linguist
Prosodie-Capgemini

Jul 2019 – Mar 2020   Madrid, Spain

- Training and error analysis of natural language understanding models
- Monitoring and improvement of the user experience
- Creation and tuning of corpus linguistics and grammars for ASR
- Installation and using Nuance System

Computational Linguist
Intercom Strategies

Dec 2018 - Jun 2019   Madrid, Spain

- Developed an automatic system for the analysis of political discourse in social networks with NLP techniques. The results are published in the journal *Expert Systems with Applications*.
- Was responsible for the processing of unstructured texts and annotated corpora.
- Trained machine learning models for text classification and clustering.

Linguist
Sigma Technologies

Jul 2016 – Nov 2018   Madrid, Spain

- Participated in projects that consisted of the evaluation of automatic speech recognition (ASR) and synthesis systems.
- Annotated training corpus according to in-home ASR guidelines.
- Was responsible for the annotation evaluation.

EDUCATION

PhD in Language Analysis and Processing
University of the Basque Country, Department of Languages and Computer Systems, Spain

Oct 2019 – March 2025

Master's Degree in Language Analysis and Processing
University of the Basque Country, Department of Languages and Computer Systems, Spain

Sep 2017 – Sep 2019

Bachelor's Degree in Journalism
Altai State University, Russia

Sep 1999 – June 2004

KNOWLEDGE & SKILLS

Machine Learning   Deep Learning
Unsupervised Learning
Information Retrieval
Information Extraction
Retrieval-Augmented Generation
Large Language Models   Data Visualization
Computational linguistics   Statistics
Scientific Writing   Project Management

FRAMEWORKS & TOOLS

Python   HF Transformers   PyTorch
Scikit Learn   LlamaIndex   Anaconda
Vector Databases   Selenium
Pandas   Docker   Elasticsearch
FAISS   Git   GitLab   Ubuntu
LaTex   RESTful API   FastAPI
Shell scripting   VSCode   SLURM

LANGUAGES

English
Spanish
Russian

PUBLICATIONS

👥 Conference Proceedings

- Platas, Alexander et al. (Sept. 2024). "Synthetic Annotated Data for Named Entity Recognition in Computed Tomography Scan Reports". In: *SEPLN-2024: 40th Conference of the Spanish Society for Natural Language Processing*. CEUR Workshop Proceedings. Valladolid, Spain: CEUR-WS.org, pp. 69–78. URL: `https://ceur-ws.org/Vol-3846/paper07.pdf`.

- Zotova, Elena, Montse Cuadros, and German Rigau (May 2024). "CliniRes: Publicly Available Mapping of Clinical Lexical Resources". In: *Proceedings of the First Workshop on Patient-Oriented Language Processing (CL4Health) @ LREC-COLING 2024*. Ed. by Dina Demner-Fushman et al. Torino, Italia: ELRA and ICCL, pp. 163–172. URL: `https://aclanthology.org/2024.cl4health-1.20`.

- Zotova, Elena (2023). "Multilingual Information Extraction in Clinical Texts Using Deep Learning Approaches". In: *Proceedings of the Doctoral Symposium on Natural Language Processing from the Proyecto ILENIA (PLN-DS-2023) held as part of the XXXIX edition of the International Conference of the Spanish Society for Natural Language Processing (SEPLN 2023), Jaén, Spain, September 28, 2023*. Ed. by María Teresa Martín-Valdivia et al. Vol. 3625. CEUR Workshop Proceedings. CEUR-WS.org, pp. 111–119. URL: `https://ceur-ws.org/Vol-3625/paper14.pdf`.

- Zotova, Elena, Montse Cuadros, and German Rigau (2023). "Towards the integration of WordNet into ClinIDMap". In: *Proceedings of the 12th Global Wordnet Conference*, pp. 352–362. URL: `https://aclanthology.org/2023.gwc-1.42.pdf`.

- Zotova, Elena, Aitor García-Pablos, Montse Cuadros, and German Rigau (2023). "VICOMTECH at MedProcNER 2023: Transformers-based Sequence-labelling and Cross-encoding for Entity Detection and Normalisation in Spanish Clinical Texts". In: *Working Notes of the Conference and Labs of the Evaluation Forum (CLEF 2023), Thessaloniki, Greece, September 18th to 21st, 2023*. Ed. by Mohammad Aliannejadi et al. Vol. 3497. CEUR Workshop Proceedings. CEUR-WS.org, pp. 206–218. URL: `https://ceur-ws.org/Vol-3497/paper-018.pdf`.

- Zotova, Elena, Montse Cuadros, and German Rigau (2022). "ClinIDMap: Towards a Clinical IDs Mapping for Data Interoperability". In: *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. Marseille, France: European Language Resources Association, pp. 3661–3669. URL: `https://aclanthology.org/2022.lrec-1.390.pdf`.

- Zotova, Elena, Aitor García-Pablos, Naiara Pérez, et al. (2022). "Vicomtech at LivingNER2022". In: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2022) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2022), A Coruña, Spain, September 20, 2022*. Vol. 3202. CEUR Workshop Proceedings. CEUR-WS.org. URL: `https://ceur-ws.org/Vol-3202/livingner-paper1.pdf`.

- Zotova, Elena, Aitor García-Pablos, and Montse Cuadros (2021). "Vicomtech at MEDDOPROF: Automatic Information Extraction and Disambiguation in Clinical Text". In: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2021) co-located with the Conference of the Spanish Society for Natural Language Processing (SEPLN 2021), XXXVII International Conference of the Spanish Society for Natural Language Processing., Málaga, Spain, September, 2021*. Vol. 2943. CEUR Workshop Proceedings. CEUR-WS.org, pp. 776–787. URL: `https://ceur-ws.org/Vol-2943/meddoprof_paper5.pdf`.

- García-Pablos, Aitor, Montse Cuadros, and Elena Zotova (2020). "Vicomtech at CAPITEL 2020: Facing Entity Recognition and Universal Dependency Parsing of Spanish News Articles with BERT Models". In: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, September 23th, 2020*. Ed. by Miguel Ángel García Cumbreras et al. Vol. 2664. CEUR Workshop Proceedings. CEUR-WS.org, pp. 52–59. URL: `https://ceur-ws.org/Vol-2664/capitel_paper3.pdf`.

- García-Pablos, Aitor, Naiara Pérez, et al. (2020). "Vicomtech at eHealth-KD Challenge 2020". In: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, September 23th, 2020*. Ed. by Miguel Ángel García-Cumbreras et al. Vol. 2664. CEUR Workshop Proceedings. CEUR-WS.org, pp. 102–111. URL: `https://ceur-ws.org/Vol-2664/eHealth-KD_paper3.pdf`.

- Zotova, Elena, Rodrigo Agerri, Manuel Nuñez, et al. (May 2020). "Multilingual Stance Detection in Tweets: The Catalonia Independence Corpus". In: *Proceedings of the Twelfth Language Resources and Evaluation Conference*. Ed. by Nicoletta Calzolari et al. Marseille, France: European Language Resources Association, pp. 1368–1375. ISBN: 979-10-95546-34-4. URL: `https://aclanthology.org/2020.lrec-1.171`.

- Zotova, Elena, Montse Cuadros, Naiara Pérez, et al. (2020). "Vicomtech at ALexS 2020: Unsupervised Complex Word Identification Based on Domain Frequency". In: *Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2020) co-located with 36th Conference of the Spanish Society for Natural Language Processing (SEPLN 2020), Málaga, Spain, September 23th, 2020*. Ed. by Miguel Ángel García-Cumbreras et al. Vol. 2664. CEUR Workshop Proceedings. CEUR-WS.org, pp. 7–14. URL: `https://ceur-ws.org/Vol-2664/alesx_paper1.pdf`.

📄 Journal Articles

- Zotova, Elena, Rodrigo Agerri, and German Rigau (2021). "Semi-automatic generation of multilingual datasets for stance detection in Twitter". In: *Expert Systems with Applications* 170, p. 114547. ISSN: 0957-4174. DOI: `https://doi.org/10.1016/j.eswa.2020.114547`. URL: `https://www.sciencedirect.com/science/article/pii/S095741742031191X`.