John Evan **Ortega**

PHD GRADUATE · COMPUTATIONAL LINGUISTICS

60 5th Avenue, New York, New York, 10011, United States

📱 (+1) 347-766-9318 | ✉ jortega@cs.nyu.edu | 🏠 johneortega.github.io | 🔘 johneortega | 🔲 johnortegajr

Academic Experience

Columbia University
New York, New York, United States

PART-TIME STAFF ASSOCIATE FACULTY - TEXT ANALYTICS
Aug 2019 - Present

- Prepare and give lectures with notes and materials for Master's level Text Analytics and Natural Understanding class.
- Give guidance, grading, and other duties related to student excellence.
- Explain complex ideas and deliver insight into real-world problems as they apply to theory in Natural Language Processing.
- www.columbia.edu

New York University
New York, New York, United States

PART-TIME NLP RESEARCHER AND INSTRUCTOR
Aug 2017 - Present

- Perform research with some of the industry's smartest people on complex Machine Learning problems.
- Publish papers in major conferences on state-of-the-art findings in various domains such as Named-Entity Recognition, Neural Networks, and Machine Translation.
- Assist with instruction in three Computer Science classes: Natural Language Processing, Statistical Natural Language Processing, and Introduction to Data Structures.
- Prepare, deliver, and grade localized planning and lessons including mid-term and final exams.
- www.nyu.edu

La Universidad de Santiago de Compostela
Santiago de Compostela, Spain

POST-DOCTORAL RESEARCHER
Sep 2021 - Present

- Work with collaborators on Proxecto Nós, a project dedicated to revitalizing the Galician language in Spain.
- Assist in funding for projects at the national and international level such as AMTEGA (municipality for Galicia), the Spanish Ministry of Education, and EU Horizon.
- Perform in-depth research into the latest state-of-the-art techniques for NLP tasks such as topic modeling, machine translation, and chatbots.
- www.usc.es

Fondazione Bruno Kessler - University of Trento, Italy
Trento, Italy

INTERN RESEARCHER
Aug 2012 - Oct 2012

- Added defined rules for co-reference relationships into the Decoref modules of the Stanford OpenNLP.
- Compared corpora against published Conll results from Stanford.
- Collaborated with the BCROCE project.
- www.fbk.eu

Professional Experience

JP Morgan Chase
New York, New York

EXECUTIVE DIRECTOR - ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING
Mar 2021 - Oct 2021

- Collaborate with all of JPMorgan's lines of business and functions to delivery software solutions.
- Experiment, develop and productionize high quality machine learning models, services and platforms to make huge technology and business impact.
- Lead a team of engineers to build products and deliver solutions, depending on previous architecture and technical leadership experience.
- Design and implement highly scalable and reliable data processing pipelines and perform analyses and insights to drive and optimize business results.
- www.jpmorganchase.com

AIG
New York, New York

SENIOR NLP RESEARCH SCIENTIST
Mar 2019 - Aug 2020

- Use Natural Language Processing and Machine Learning to help determine risk for various insurance applications at tech arm called Blackboard.
- Advise junior team members and lead the research department to textual methodology solutions with modeling in various programming languages like Python, Java, R, and C++.
- www.aig.com

MAY 21, 2022   JOHN E. ORTEGA · CV-RESUME   1

Nuance Communications
New York, New York

Senior NLP Research Scientist
May 2012 - Apr 2019

- Perform research and create applications using various programming languages such as Java, UIMA-AS, C++, PHP, Python, Perl, and Bash.
- Implement experiments with complex neural networks using tools like Theano, TensorFlow, Keras, and Spark in a GPU environment.
- Work within a remote Linux environment using Clinical Language Understanding and Natural Language Processing.
- Create and develop accuracy and precision reporting tools through analysis of rules-based and unsupervised learning techniques for Computer Assisted Coding and other language models.
- Advise and implement software required for novel approaches in the clinical setting.
- Design, develop, and manage Amazon Web Service health applications on CentOS.
- Implement multiple tier systems in Java for algorithmic language semantics on Linux and Windows.
- www.nuance.com

Education

University of Alicante
Alicante, Spain

Ph.D. in Computer Science
Oct 2013 - Mar 2021

- Focus on natural language processing with a thesis on machine translation for computer-assisted translation tools.
- Main advisors: Mikel Forcada and Felipe Sánchez Martínez
- International Mention with a stay at New York University.
- www.ua.es

Hofstra University
Hempstead, New York

M.S. in Computer Science
Aug 2011 - May 2013

- Focus on natural language processing with a thesis on low-resource machine translation.
- Main advisor: Krishnan Pillaipakkamnatt
- Achieved Cum Laude with a perfect 4.0 score for all classes.
- www.hofstra.edu

Augusta State University
Augusta, Georgia

B.S. in Computer Science
Aug 1998 - Dec 2001

- Focus on software engineering and databases.
- Main advisor: Robert Jarman
- www.augusta.edu

Program Committees

2022	**Reviewer**, EURALI 2022 : Wkshp. on Res. and Tech. for Indig., Endang. and Lesser resource Langs. in Eurasia	*Virtual*
2022	**Organizer**, IWSLT 2022: International Workshop on Spoken Language Translation (low-resources)	*Virtual*
2022	**Reviewer**, NeTTT 2022: Conference on New Trends in Translation and Technology	*Virtual*
2022	**Organizer and Reviewer**, ALTNLP 2022 : 1st Int. Conf. on Agglutinative Langs. as a Challenge of NLP	*Virtual*
2022	**Reviewer**, ACL 2022: 60th Annual Meeting of the Association for Computational Linguistics	*Virtual Dublin*
2022	**Reviewer**, PROPOR 2022: 15th International Conference on Computational Processing of Portuguese	*Virtual*
2022	**Reviewer**, INSIGHTS 2022: Workshop on Insights from Negative Results in NLP	*Virtual ACL*
2022	**Reviewer**, LREC 2022: 13th Edition of its Language Resources and Evaluation Conference	*Virtual*
2021	**Moderator and reviewer**, MT Summit 2021: 18th Machine Translation Summit	*Virtual*
2021	**Organizer and Reviewer**, AmericasNLP 2021: 1st WS on NLP for Indigenous Langs. of the Americas	*Virtual EMLP*
2021	**Main Chair and Organizer**, LoRESMT 2021: 4th Workshop on Tech. for MT of Low Resource Languages	*Virtual*
2021	**Reviewer**, Springer Journal: Special Topics on Machine Translation	*Virtual*
2021	**Reviewer**, Springer Journal: Language Resources and Evaluation	*Virtual*
2021	**Reviewer**, MDPI Journal: Informatics Publications on Fuzzy-Match Repair	*Virtual*
2021	**Moderator and Reviewer**, ACL 2021: 59th Annual Meeting of the Association for Computational Linguistics	*Virtual*
2020	**Moderator and Reviewer**, AACL-IJCNLP 2020: 1st Conference of the Asia-Pacific Chapter of the ACL	*Virtual*
2020	**Main Chair and Organizer**, LoRESMT 2020: 3rd Workshop on Tech. for MT of Low Resource Languages	*Virtual*
2020	**Moderator**, AMTA 2020: 15th biennial conference of the Association for Machine Translation	*Virtual*
2020	**Main Chair and Organizer**, PEMDT 2020: 1st Workshop on Post-Editing in Modern-Day Translation	*Virtual*

Presentations

Big Data Conference Europe
Vilnius, Lithuania

How Can Artificial Intelligence use Big Data for Translating Documents?
Nov 2019

- Introduced machine translation concepts to an international crowd.
- Presented several machine learning techniques for translating on big data.
- Click for video on **www.youtube.com**

Big Data Spain *Madrid, Spain*

Using Big Data to Assist with Medical Report Coding *Nov 2018*

- Covered several concept on NLP for the clinical domain.
- Showed different approaches for machine learning on healthcare problems.
- Click for video on **www.youtube.com**

AI Ukraine *Kyiv, Ukraine*

Using Big Data to Assist with Low Resources in Natural Language Processing *Sep 2019*

- Explained how the lack of resources works in NLP.
- Developed several approaches for MT and NLP in low resource settings.
- Click for video on **www.youtube.com**

Publications

Journal Articles

Revisiting CCNet for Quality Measurements in Galician
    John E. Ortega, Iria de-Dios-Flores, José Ramom Pichel, Pablo Gamallo
Computational Processing of the Portuguese Language, 2022

Fuzzy-match repair guided by quality estimation
    John E Ortega, Mikel L Forcada, Felipe Sanchez-Martinez
IEEE Transactions on Pattern Analysis and Machine Intelligence (2020). IEEE, 2020

Neural machine translation with a polysynthetic low resource language
    John E Ortega, Richard Castro Mamani, Kyunghyun Cho
Machine Translation 34.4 (2020) pp. 325–346. Springer, 2020

The Termolator: terminology recognition based on chunking, statistical and search-based scores
    Adam Laurence Meyers, Yifan He, Zachary Glass, John Ortega, Shasha Liao, Angus Grieve-Smith, Ralph Grishman, Olga Babko-Malaya
Frontiers in Research Metrics and Analytics 3 (2018) p. 19. Frontiers, 2018

Patents

Computer assisted coding systems and methods
    Oscar Ferrandez-Escamez, Neil D Barrett, Ravi Kondadadi, Girija Yegnanarayanan, Brian William Delaney, John Ortega
US Patent App. 15/632,152, 2018

Medical report coding with acronym/abbreviation disambiguation
    Ravi Kondadadi, Girija Yegnanarayanan, Brian William Delaney, John Ortega
US Patent App. 15/045,167, 2017

Conference Proceedings

Findings of the IWSLT 2022 Evaluation Campaign
    Antonios Anastasopoulos, Loïc Barrault, Luisa Bentivogli, Marcely Zanon Boito, Ondřej Bojar, Roldano Cattoni, Anna Currey, Georgiana Dinu, Kevin Duh, Maha Elbayad
Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022), 2022

ON-TRAC Consortium Systems for the IWSLT 2022 Dialect and Low-resource Speech Translation Tasks
    Marcely Zanon Boito, John Ortega, Hugo Riguidel, Antoine Laurent, Loïc Barrault, Fethi Bougares, Firas Chaabani, Ha Nguyen, Florentin Barbier, Souhir Gahbiche
Proceedings of the 19th International Conference on Spoken Language Translation (IWSLT 2022), 2022

AmericasNLI: Evaluating Zero-shot Natural Language Understanding of Pretrained Multilingual Models in Truly Low-resource Languages
    Abteen Ebrahimi, Manuel Mager, Arturo Oncevay, Vishrav Chaudhary, Luis Chiruzzo, Angela Fan, John Ortega, Ricardo Ramos, Annette Rios, Ivan Vladimir
arXiv preprint arXiv:2104.08726 (2021). 2021

Findings of the AmericasNLP 2021 Shared Task on Open Machine Translation for Indigenous Languages of the Americas
    Manuel Mager, Arturo Oncevay, Abteen Ebrahimi, John Ortega, Annette Rios Gonzales, Angela Fan, Ximena Gutierrez-Vasques, Luis Chiruzzo, Gustavo Giménez Lugo, Ricardo Ramos
Proceedings of the First Workshop on Natural Language Processing for Indigenous Languages of the Americas, 2021

Findings of the LoResMT 2021 Shared Task on COVID and Sign Language for Low-resource Languages
    Atul Kr. Ojha, Chao-Hong Liu, Katharina Kann, John Ortega, Sheetal Shatam, Theodorus Fransen
Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021), 2021, Virtual

Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021)
    John Ortega, Atul Kr. Ojha, Katharina Kann, Chao-Hong Liu
Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021), 2021, Virtual

Love Thy Neighbor: Combining Two Neighboring Low-Resource Languages for Translation
    John E. Ortega, Richard Alexander Castro Mamani, Jaime Rafael Montoya Samame
Proceedings of the 4th Workshop on Technologies for MT of Low Resource Languages (LoResMT2021), 2021, Virtual

Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages

Alina Karakanta, Atul Kr Ojha, Chao-Hong Liu, Jade Abbott, John Ortega, Jonathan Washington, Nathaniel Oco, Surafel Melaku Lakew, Tommi A Pirinen, Valentin Malykh

Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages, 2020

Overcoming Resistance: The Normalization of an Amazonian Tribal Language
John Ortega, Richard Alexander Castro-Mamani, Jaime Rafael Montoya Samame

Proceedings of the 3rd Workshop on Technologies for MT of Low Resource Languages, 2020

Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation
John E. Ortega, Marcello Federico, Constantin Orasan, Maja Popovic

Proceedings of 1st Workshop on Post-Editing in Modern-Day Translation, 2020, Virtual

Improving Translations by Combining Fuzzy-Match Repair with Automatic Post-Editing
John Ortega, Felipe Sánchez-Martínez, Marco Turchi, Matteo Negri

Proceedings of Machine Translation Summit XVII Volume 1: Research Track, 2019, Dublin, Ireland

A Comparison of Machine Translation Paradigms for Use in Black-Box Fuzzy-Match Repair
Rebecca Knowles, John Ortega, Philipp Koehn

Proceedings of the AMTA 2018 Workshop on Translation Quality Estimation and Automatic Post-Editing, 2018

Using Morphemes from Agglutinative Languages like Quechua and Finnish to Aid in Low-Resource Translation
John Ortega, Krishnan Pillaipakkamnatt

Proceedings of the AMTA 2018 Workshop on Technologies for MT of Low Resource Languages (LoResMT 2018), 2018

Letting a neural network decide which machine translation system to use for black-box fuzzy-match repair
John Evan Ortega, Weiyi Lu, Adam Meyers, Kyungyun Cho

Proceedings of the 21st Annual Conference of the European Association for Machine Translation: 28-30 May 2018, Universitat d'Alacant, Alacant, Spain, 2018

A comparative study of classifying legal documents with neural networks
Samir Undavia, Adam Meyers, John E Ortega

2018 Federated Conference on Computer Science and Information Systems (FedCSIS), 2018

Fuzzy-match repair using black-box machine translation systems: what can be expected?
John Ortega, Felipe Sánchez-Martínez, Mikel L Forcada

Conferences of the Association for Machine Translation in the Americas: MT Researchers' Track, 2016

Using any machine translation source for fuzzy-match repair in a computer-aided translation setting
John E Ortega, Felipe Sánchez-Martínez, Mikel L Forcada

Proceedings of the 11th Conference of the Association for Machine Translation in the Americas: MT Researchers Track, 2014