

Madrazo Azpiazu, Ion

PHD IN COMPUTER SCIENCE

✉ ion.madrazo.azpiazu[at]gmail.com | 🏠 ionmadrazo.github.io | 📧 ionmadrazo | 📺 ionmadrazo

Summary

PhD in Computer Science (expected Fall 2019) with an emphasis in Natural Language Processing, Information Retrieval, and Deep Learning. Publications on top-tier conferences and journals including TACL, JASIST, ACM SIGIR, ACM RecSys, and ACM CHIIR. Born and grown in multicultural environments with knowledge of 4 languages.

Work Experience

Zalando Research

DATA SCIENCE INTERN

- Research and development of an automatic tool for determining the fit of jeans given catalogue images.

Dublin, Ireland

Summer 2018

Boise State University

RESEARCH ASSISTANT

- Research and development of Natural Language Processing and Information Retrieval based tools tailored to the educational domain.

Boise, U.S.A

2015 – 2019

Boise State University

TEACHER ASSISTANT

- Lead instructor of Computer Science I Lab (1 Semester)
- Teacher assistant of Information Retrieval Course (4 Semesters, guest lecturer at least twice per semester)
- Teacher assistant of Artificial Intelligence Course (1 Semester)
- Teacher assistant of Computer Theory Course (4 Semesters)
- Teacher assistant of Computer Science II (1 Semester)

Boise, U.S.A

2015 – 2019

IXA NLP research group

RESEARCH ASSISTANT

- Research and development of natural language processing applications using data mining and linguistic tools.

Donostia, Spain

2012 – 2014

Publicidad en almibar

WEB DEVELOPER

- Design, development, and testing of webpages and web applications.

Azkoitia, Spain

2010 – 2014

Education

Boise State University

PHD IN COMPUTER SCIENCE

Boise, U.S.A

2017 – 2019

University of Deusto, Rovira i Virgili University

INTERNATIONAL SUMMER SCHOOL ON DEEP LEARNING 2017

Bilbao, Spain

Summer 2017

Boise State University

MASTER'S DEGREE: COMPUTER SCIENCE

- GPA 4.0/4.0.

Boise, U.S.A

2015 – 2016

University of Basque Country

MASTER'S DEGREE: COMPUTATION AND INTELLIGENT SYSTEMS

- Average score: 9.2/10.

Donostia

2013 – 2014

University of Basque Country

BACHELOR'S DEGREE: COMPUTER SCIENCE, COMPUTATION SPECIALIZATION

- Average score: 8.45/10
- Top 1 student of specialization
- Score of 10/10 with High Honors in the end of degree project

Donostia

2009 – 2013

Awards & Grants

2015-19	Grant , Boise State Graduate Assistantship	<i>Boise, U.S.A</i>
2018	Award , Boise State Outstanding Graduate Student	<i>Boise, U.S.A</i>
2018	Grant , ACM CHI Student Travel Grant	<i>Boise, U.S.A</i>
2017	Grant , ACM SIG Student Travel Grant	<i>Boise, U.S.A</i>
2013	Award , Special award for academic performance in Bachelor's degree	<i>Donostia, Spain</i>
2013	Grant , Univesity of Basque Country Grant for Master's Degree	<i>Donostia, Spain</i>
2012	Award , Entreprenari award for best end of degree project with entrepreneurial nature	<i>Donostia, Spain</i>

Technical Skills

Legend: Used in [most projects](#) / [several projects](#) / [one or two projects](#).

Machine/Deep Learning	Pytorch , Tensorflow , Numpy , Sklearn
Data processing & storage	Pandas , SQL , Elastic Search , Lucene , Hadoop (HDFS, Mapreduce) , Spark
NLP Toolkits	NLTK , Spacy , CoreNLP , Gensim
Visualization	MatPlotLib , Seaborn , GGplot2 , Tableau
Clusters & containerization	Slurm , Docker
Programming Languages	Python , Java , R , Bash , C , C++ , C# , Perl , Haskell
Web & Mobile Dev	Symfony 2 , Bootstrap , Jquery , Jquery UI , AngularJS , Android SDK
Architectures	LSTM , GRU , Attention models , Transformers , Residual Networks
Text Representation	(Cross-lingual) Word-Embeddings , TF-IDF , Bag-of-words
Recommendation Systems	Item-Item , User-User , Matrix Factorization , Association Rules
Evaluation	(Paired) T-test , Cross-fold-validation , A/B testing
Pretrained models	FastText , Word2Vec , BERT , ImageNet , GloVe
Other Algorithms	Stochastic Gradient Descent , PCA , SVD , Boosting , Genetic Algorithms
Others	Git , L^AT_EX , Arduino

Publications

JOURNALS

- **Ion Madrazo Azpiazu**, Maria Soledad Pera, Multiattentive hierarchical recurrent neural network architecture for automatic readability assessment. Transactions of the Association for Computational Linguistics (TACL), Transactions of the Association for Computational Linguistics (Vol. 7 pp. 421-436), 2019.
- **Ion Madrazo Azpiazu**, Maria Soledad Pera, Is Multilingual Readability Assessment Possible? Journal of the Association for Information Science and Technology (JASIST), to appear.
- **Ion Madrazo Azpiazu**, Nevena Dragovic, Maria Soledad Pera, Jerry Alan Fails, Online searching and learning: YUM and other search tools for children and teachers. Information Retrieval Journal (Vol. 20, No. 5, pp. 524-545), 2017.

CONFERENCE PROCEEDINGS

- **Ion Madrazo Azpiazu**, Nevena Dragovic, Oghenemaro Anuyah, Maria Soledad Pera. Looking for the Movie Seven or Sven from the Movie Frozen? A Multi-perspective Strategy for recommending Queries for Children. Proceedings of the Conference on Human Information Interaction and Retrieval (ACM CHIIR), pages 92–101, 2018.
- Michael D. Ekstrand, Mucun Tian, **Ion Madrazo Azpiazu**, Jennifer D. Ekstrand, Oghenemaro Anuyah, David McNeill, Maria Soledad Pera. All The Cool Kids, How Do They Fit In? Popularity and Demographic Biases in Recommender Evaluation and Effectiveness. Conference on Fairness, Accountability, and Transparency (FAT), pages 172–186, 2017.
- Hoda Mehrpouyan, **Ion Madrazo Azpiazu**, and Maria Soledad Pera. Measuring Personality for Automatic Elicitation of Privacy Preferences. IEEE Symposium on Privacy-Aware Computing (PAC), pages 84–95, 2017.
- Nevena Dragovic, **Ion Madrazo Azpiazu**, and Maria Soledad Pera. "Is Sven Seven?": A Search Intent Module for Children". International ACM Special Interest Group on Information Retrieval Conference (ACM SIGIR), pages 885–888, 2016.
- Itziar Aldabe Arregi, Itziar González Dios, Inigo López Gazpio, **Ion Madrazo Azpiazu**, and Montse Maritxalar Anglada. Two Approaches to Generate Questions in Basque. In the Spanish Society for Language Processing (SEPLN 2013),

WORKSHOPS, SYMPOSIUMS, AND OTHER CONTRIBUTIONS (PEER-REVIEWED)

- Oghenemaro Anuyah, **Ion Madrazo Azpiazu** and Maria Soledad Pera. Lightning Talk - Looking for the Movie Seven or Sven from the Movie Frozen?: A multi-perspective strategy for recommending queries for children. In: Companion proceedings of World Wide Web Conference (Track: Women in Web Data Science), pages 1266–1267, 2019.
- **Ion Madrazo Azpiazu**, Michael Green, Oghenemaro Anuyah, and Maria Soledad Pera. Can we leverage rating patterns from traditional users to enhance recommendations for children?. In: Poster Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18), Vancouver, Canada, 2018.
- Michael Ekstrand, **Ion Madrazo Azpiazu**, Katherine Landau Wright, and Maria Soledad Pera. Retrieving and Recommending for the Classroom: Stakeholders, Objectives, Resources, and Users. In: Proceedings of the 2nd Workshop on Recommendation in Complex Scenarios; co-located with ACM Conference on Recommender Systems, Vancouver, Canada, 2018.
- Nevena Dragovic, **Ion Madrazo Azpiazu**, and Maria Soledad Pera. From Recommendation to Curation: When the system becomes your personal docent. In: Joint Workshop on Interfaces and Human Decision Making for Recommender Systems (IntRS); co-located with ACM Conference on Recommender Systems, Vancouver, Canada, 2018.
- Oghenemaro Anuyah, **Ion Madrazo Azpiazu**, Maria Soledad Pera. Can Readability Enhance Recommendations on Community Question Answering Sites? In: Poster Proceedings of the 11th ACM Conference on Recommender Systems (RecSys '17), Lake Como, Italy, 2017.
- **Ion Madrazo Azpiazu**, Guadalupe Guereta, Santiago Diez, Martin Bianculli Martin, Alejandro Gonzalez, Carlos Flury, Franco Ferrari, Mariano Baudena, Guillermo Leale, and Maria Soledad Pera. Comparative analysis on text distance measures applied to Community Question Answering data. In: Congreso Nacional de Ingenieria Informatica (CONAISI), 2017.
- **Ion Madrazo Azpiazu** and Maria Soledad Pera. Is Readability a Valuable Signal for Hashtag Recommendations?. In: Poster Proceedings of the 10th ACM Conference on Recommender Systems (RecSys '16), Boston, USA, 2016.
- **Ion Madrazo Azpiazu**, Nevena Dragovic, and Maria Soledad Pera. Finding, Understanding and Learning: Making Information Discovery Tasks Useful for Children and Teachers. In: Workshop on Search as Learning (SAL); co-located with ACM SIGIR conference, Pisa, Italy, 2016.

Main Projects

Hierarchical cross-lingual word embeddings

Ongoing

PYTORCH, WORD EMBEDDINGS, UNSUPERVISED LEARNING TECHNIQUES, CROSS-LINGUAL REPRESENTATIONS

Boise State University

- Word embeddings have drastically changed how we build text processing applications, given their unprecedented capabilities for representing the semantics of words. Traditional word embedding generation strategies, however, generate different embeddings for the same word depending on the language they are trained for. In this work we do research on cross-lingual word embedding generation strategies. Despite existing techniques, the strategy we are creating does not rely on a pivot language (such as English) to generate the cross-lingual space, but takes advantage of the inherent hierarchical nature of world-languages to construct such space. This hierarchical representation allows better incorporating languages that are low-resource and typologically distant from English.

Automatically determining the fit of jeans

Summer 2018

IMAGENET, TENSORFLOW, TABLEAU, CONVOLUTIONAL NEURAL NETWORKS, BRAND AND PRODUCT EMBEDDINGS

Zalando

- As many online shops, Zalando assigns a significant amount of its expenses to transportation costs, which include costs for product delivery to the customers, along with product returns (when necessary). Therefore, reducing transportation costs could drastically boost company income. This can be done by reducing the amount of returned products. Jeans are one of the most frequently-returned products, given that their fit/size labelling can be misleading. During my internship at Zalando, I designed and developed an automatic classification strategy for determining the fit of jeans based on their catalogue image. This tool can provide suggestions such as "this pair of jeans might run tighter/smaller than usual". Based on the developed tool, I conducted a study and produced a report enumerating brands that have jeans and the corresponding size/fit labelling patterns, e.g., "this brand tends to produce jeans that are tighter than what they label".

Multiattentive hierarchical recurrent neural network for readability assessment

2018-2019

ATTENTION MECHANISMS, RECURRENT NEURAL NETWORKS, CROSS-LINGUAL TEXT CLASSIFICATION, WORD EMBEDDINGS

Boise State University

- During this project, we proposed a neural network architecture that involves recurrent layers and multiple attention mechanisms for predicting the readability level of a document while also presenting the user with explanations on which parts of the document are harder or simpler and why. The architecture is designed to fit the characteristics of different languages. As a result, the first prototype can work in 6 languages. Different from CRAS (see below), this strategy does not depend on any human-engineered feature, making it easily extensible to any language that has leveled corpora.

CRAS: Cross-lingual Readability Assessment System

2016-2018

TOKENIZATION, LEMMATIZATION, PART OF SPEECH TAGGING, NAMED ENTITIES, DEPENDENCY PARSING, WORDNET

Boise State University

- CRAS is a cross-lingual readability system. It is based on a supervised learning approach capable of learning features for a number of languages at a time. All the features considered are designed to independently of the input language, enabling CRAS to predict readability of languages that have barely learned about, using knowledge acquired from other languages.

YouUnderstood.me? Readability Filtered Search Environment for Children

2016-2018

LEARNING TO RANK, FILTERING, LEMMATIZATION, PART OF SPEECH TAGGING, TF-IDF, WORDNET, LUCENE, CORENLP

Boise State University

- YouUnderstood.me (YUM) is an online environment oriented to make information discovery tasks valuable for children. YUM can be seen as an intermediate layer between a search engine and a child, that tries to amend issues children encounter while using popular search engines. YUM takes advantage of QuK and ReQuK (a search intent strategy and a query suggestion strategy we developed, respectively) as well as a readability based filtering strategy and a tracking systems that permits a teacher to follow the progress of his students in terms of reading skills.

ReQuK: Recommendations Based on Query Intent for Kids

2016-2018

NEURAL NETWORKS, LSTM, WIDE AND DEEP ARCHITECTURE, WORD EMBEDDINGS, TOKENIZATION, LEMMATIZATION

Boise State University

- ReQuK was created based on QuK with the purpose of providing query recommendation that children would find relevant. It is based on a multi-criteria recommendation system that considers patterns in queries that may be of children's interest, in order to recommend queries containing them.

Automatic Elicitation of Privacy Preferences Based on Personality Traits

2016-2017

RANDOM FORESTS, CORRELATION, DATASET BALANCING, MATPLOTLIB, GGLOT, PANDAS, NUMPY, SKLEARN

Boise State University

- In this project we analyzed personality traits and generated patterns to predict whether a person should be willing to share certain types of information or not. We also analyzed differences in these patterns depending on who people are sharing information with and how this was related to their personality.

QuK: Query Intent for Kids

2015-2017

SPELL CHECKING, TF-IDF, WORDNET, LUCENE, CORENLP

Boise State University

- QuK is a search intent detection tool for children, capable of capturing the main intent of a query from a children perspective, and provide a keyword based reformulated query that search engines are better in understanding. For doing so, QuK considers common misspelling patterns among children and vocabulary related to them such as children TV Shows entities, among other features.

ARASB: Automatic Readability Assessment for Basque

2014

REGRESSION, COST SENSITIVE LEARNING, BAYESIAN LEARNING, WORDNET, FEATURE SELECTION, DEPENDENCY PARSING

University of the Basque Country

- ARASB is an automatic readability assessment tool specifically designed for Basque. It is based on a supervised learning approach, that considers lexical, syntactic, semantic and morphological features. ARASB was the the second existing readability assessment tool for Basque and the first one to consider syntactic and semantic features.

QG-Maltixa: Automatic question generation for Basque

2013

TOKENIZING, LEMMATIZING, PART OF SPEECH TAGGING, DEPENDENCY PARSING, MUTUAL INFORMATION, ENTROPY

University of the Basque Country

- QG-Maltixa is an automatic question generation tool. It was conceived as a way for children to work on comprehension of texts without the need of a teacher to guide them. QG-Maltixa is capable of generating factoid question/answer pairs that are given to the student so that he can test content comprehension on his own.

Languages

English	Proficient
Spanish	Native
Basque	Native
German	Lower-intermediate