



Marvin Matías Agüero Torales

Nombres: Marvin Matías

Sr. - Nacionalidad Paraguaya

Apellidos: Agüero Torales

Nombre en citas bibliográficas: Marvin M. Agüero-Torales o Marvin Agüero-Torales, M. M. Agüero-Torales, Marvin Agüero

Nacido el 04-05-1987 en Áregua

Sin datos adicionales

marvin-aguero@outlook.com

Datos del PRONII

Área: **Ingeniería y Tecnología - Asociado**
 Categorización Actual: **Nivel I - Res.: 570/2022**
 Ingreso al PRONII: **Nivel I - Res.: 570/2022**

Información de Contacto

Mail: **marvin-aguero@outlook.com**
 Pagina Web: **https://www.researchgate.net/profile/Marvin_Aguero-Torales**
 Pagina Web: **<https://www.linkedin.com/in/marvin-aguero/>**
 Pagina Web: **<https://github.com/mmaguero>**
 Pagina Web: **<https://scholar.google.com/citations?user=JqrU3FUAAAAJ&hl;>**
 Pagina Web: **<https://orcid.org/0000-0002-0910-0310>**
 Pagina Web: **[https://zenodo.org/search?page=1&size=20&q="](https://zenodo.org/search?page=1&size=20&q=)** Agüero, Marvin"
 Pagina Web: **[https://dataverse.harvard.edu/dataverse/harvard/?q="](https://dataverse.harvard.edu/dataverse/harvard/?q=)** Agüero-Torales"
 Pagina Web: **<https://twitter.com/marvinaguerot>**
 Pagina Web: **<https://granada.academia.edu/mmaguero>**
 Pagina Web: **<https://www.kaggle.com/mmaguero>**
 Pagina Web: **<https://huggingface.co/mmaguero>**
 Pagina Web: **tinyurl.com/simpleCVMarvinAguero**

Áreas de Actuación

- 1 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Información y Bioinformática, Minería de Texto
- 2 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Computación, Ingeniería y modelado de datos
- 3 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Información y Bioinformática, Procesamiento de Lenguaje Natural
- 4 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Información y Bioinformática, Inteligencia Artificial
- 5 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Información y Bioinformática, Ciencias de datos
- 6 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Información y Bioinformática, Aprendizaje automático
- 7 Humanidades, Lengua y Literatura, Lingüística, Lingüística Computacional
- 8 Ciencias Naturales, Ciencias de la Computación e Información, Ciencias de la Información y Bioinformática, Grandes Modelos de Lenguaje

Formación Académica

DOCTORADO: Programa de Doctorado en Tecnologías de la Información y la Comunicación (2017-2022)

Universidad de Granada, España

Título: Machine Learning approaches for Topic and Sentiment Analysis in multilingual opinions and low-resource languages: From English to Guarani

Año de Obtención: 2022

Tutor: Antonio Gabriel Lopez Herrera

Sitio web de la tesis/disertación: <https://digibug.ugr.es/handle/10481/72863>

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Recuperación de información, web e internet
 Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Procesamiento de Lenguaje Natural
 Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Inteligencia Artificial
 Humanidades / Lengua y Literatura / Lingüística / Lingüística Computacional
 Ciencias Naturales / Matemáticas / Matemática Aplicada / Aprendizaje automático

MAESTRÍA: Máster Universitario en Ingeniería Informática (2016-2017)

Universidad de Granada, España

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Tecnologías InformáticasCiencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ciencia de datos

GRADO: Ingeniería en Sistemas Informáticos (2011-2013)

Universidad Autonoma de Asunción, Paraguay

Título: Sistemas de bases de datos federadas para la gestión de la información

Año de Obtención: 2015

Tutor: Mario Ruben Canto

Sitio web de la tesis/disertación: <http://dx.doi.org/10.13140/RG.2.2.29045.88802>

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Sistemas InformáticosCiencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Sistemas Informáticos

GRADO: Licenciatura en Ciencias Informáticas, esp. en Sistemas Informáticos (2006-2010)

Universidad Autonoma de Asunción, Paraguay

Título: Exámenes de suficiencia finales de carrera

Año de Obtención: 2011

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Sistemas Informáticos

Formación Complementaria

TALLERES: Iberian Languages Evaluation Forum (2024)

Sociedad Española para el Procesamiento del Lenguaje Natural , España

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

TALLERES:

Computational Approaches to Linguistic Code-Switching, CALCS 2021 (2021)

Association for Computational Linguistics, Estados Unidos

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Computational LinguisticsCiencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / natural language processing

TALLERES: Regional Machine Learning Workshop Galicia 2019 (2019)

Universidad de la Coruña, España

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Machine Learning

TALLERES: TALLER DE VISUALIZACIÓN DE DATOS (2018)

Universidad Comunera, Paraguay

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / DATA ANALYSIS

TALLERES: TALLER BI - Business intelligence (2018)

Universidad Comunera, Paraguay

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Business intelligence

POSTDOCTORADO: (2023-En Marcha)

Fujitsu, España

Título: LEIA: Lengua Española e Inteligencia Artificial

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de lenguaje natural

POSTDOCTORADO: (2022-2023)

Fujitsu, España

Título: Medigenomics

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

CURSOS DE CORTA DURACIÓN: (2024-2024)

Databricks, Estados Unidos

Título: Generative AI Fundamentals Generative AI Fundamentals

Horas totales: 30

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

CURSOS DE CORTA DURACIÓN: (2024-2025)

University of California, Berkeley, Estados Unidos

Título: Large Language Model Agents MOOC - Berkeley RDI

Horas totales: 320

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

CURSOS DE CORTA DURACIÓN: (2023-2024)

Microsoft Learn, Estados Unidos

Título: Microsoft Certified: Azure AI Engineer Associate

Horas totales: 320

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA

CURSOS DE CORTA DURACIÓN: (2022-2023)

DeepLearning.AI, Estados Unidos

Título: Natural Language Processing with Classification and Vector Spaces

Horas totales: 320

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

CURSOS DE CORTA DURACIÓN: (2021-2021)

Asociación Mexicana para el Procesamiento del Lenguaje Natural (Mexican Association for Natural Language Processing), México

Título: Mexican NATURAL LANGUAGE PROCESSING Summer School

Horas totales: 30

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSING

CURSOS DE CORTA DURACIÓN: (2020-2020)

CERTIPROF, Estados Unidos

Título: CertiProf Scrum Foundation Professional Certificate (SFPC)

Horas totales: 30

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación /

CURSOS DE CORTA DURACIÓN: (2019-2019)

Universidad de la Coruña, España

Título: Natural Language Processing

Horas totales: 6

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Natural Language Processing

CURSOS DE CORTA DURACIÓN: (2018-2018)

Universidad del Cono Sur de las Américas, Paraguay

Título: Java Web (Spring Framework + Angular JS)

Horas totales: 72

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Desarrollo Web

CURSOS DE CORTA DURACIÓN: (2018-2018)

University of Pennsylvania, Estados Unidos

Título: Gamification

Horas totales: 50

Áreas de Conocimiento:

Ciencias Sociales / Psicología / Psicología / Gamification Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / UX, UI, arquitectura de sistemas

CURSOS DE CORTA DURACIÓN: (2015-2015)

Universidad Católica "Ntra. Sra. de la Asunción", Paraguay

Título: Android Application Development

Horas totales: 96

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Programación

CURSOS DE CORTA DURACIÓN: (2015-2015)

Universidad Católica "Ntra. Sra. de la Asunción", Paraguay

Título: Java Fundamentals for Android Development

Horas totales: 12

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Programación

CURSOS DE CORTA DURACIÓN: (2014-2014)

Excelsis SACIG, Paraguay

Título: Oracle Database 11g: SQL Tuning Workshop

Horas totales: 36

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Programación

CURSOS DE CORTA DURACIÓN: (2014-2014)

Excelsis SACIG, Paraguay

Título: Oracle Reports Developer 10g: Build Reports

Horas totales: 36

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Programación

CURSOS DE CORTA DURACIÓN: (2013-2013)

Excelsis SACIG, Paraguay

Título: Oracle Forms Developer 10g: Build Internet Applications

Horas totales: 36

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Programación

CURSOS DE CORTA DURACIÓN: (2012-2012)

INFOCENTER, Paraguay

Título: Programador en SAP Portal - Web Dynpro Java Nivel Inicial

Horas totales: 96

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / ERP

CURSOS DE CORTA DURACIÓN: (2012-2012)

INFOCENTER, Paraguay

Título: Introducción al Sistema Informático SAP R3

Horas totales: 36

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / ERP

CONGRESOS: XL Congreso Internacional de la Sociedad Española para el Procesamiento del Lenguaje Natural. (2024)

Sociedad Española para el Procesamiento del Lenguaje Natural , España

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / PROCESAMIENTO DE LENGUAJE NATURAL Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / COMPUTACION LINGUISTICA

CONGRESOS: 2021 Annual Conference of the North American Chapter of the ACL (2021)

Association for Computational Linguistics, Estados Unidos

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSING / Humanidades / Lengua y Literatura / Lingüística / Computational Linguistics

CONGRESOS: CONGRESO HUMANIDADES DIGITALES HISPANICAS 2021 (2021)

HUMANIDADES DIGITALES HISPANICAS, España

Áreas de Conocimiento:

Humanidades / Otras Humanidades / HUMANIDADES DIGITALES

CONGRESOS: 7th ITQM Information technology and quantitative management based on Artificial Intelligence (2019)

Universidad de Granada, España

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática /

CONGRESOS: Big Data Coruña 2019 (2019)

Universidad de la Coruña, España

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática /

CONGRESOS: I Workshop Ciencias de Datos (2018)

Universidad Comunera, Paraguay

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ciencia de datos

CONGRESOS: CLEI 2010 (2010)

Facultad Politecnica Universidad Nacional de Asuncion, Paraguay

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Telecomunicaciones /

CONGRESOS: JIT-CITA (2009)

Universidad Autonoma de Asunción, Paraguay

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Telecomunicaciones /

CONGRESOS: JIT-CITA (2008)

Universidad Autonoma de Asunción, Paraguay

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Telecomunicaciones /

CONGRESOS: JIT-CITA (2007)

Universidad Autonoma de Asunción, Paraguay

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Telecomunicaciones /

CONGRESOS: JIT-CITA (2006)

Universidad Autonoma de Asunción, Paraguay

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Telecomunicaciones /

Idiomas

Catalán	Lee: regular	Habla: no	Escribe: no	Comprende: regular
Español	Lee: muy bien	Habla: muy bien	Escribe: muy bien	Comprende: muy bien
Gallego	Lee: regular	Habla: no	Escribe: no	Comprende: regular
Guaraní	Lee: muy bien	Habla: muy bien	Escribe: muy bien	Comprende: muy bien
Inglés	Lee: muy bien	Habla: regular	Escribe: muy bien	Comprende: muy bien
Portugués	Lee: regular	Habla: no	Escribe: no	Comprende: regular

Institución principal donde desarrolla sus actividades

Fujitsu - Fujitsu

Experiencia Profesional

Analista de Sistemas Informáticos Senior - Canales Digitales (2013 - 2018)

BASA - Banco Atlas S.A.

Carga Horaria: 42

Régimen: Dedicación total 42

Otras Informaciones:

Segundo período: 10/2017 a 06/2018

Tareas, como sénior, agregué (y mejoré) mis habilidades en:

* Canales digitales, canales alternativos y medios de pago / Nuevos productos y servicios / Metodologías ágiles / UI / UX / Análisis Funcional y Técnico / DevOps / Desarrollo Full-Stack / Microservicios.

Conocimientos adquiridos/aplicados

* Gamificación.

* Intranets, extranets, web services, servlets, API, REST, middleware, banca web, banca móvil, medios de pago; SQL scripting, Core "IT Gestión Financiera" (<http://www.bbsweb.com.ar/it-gestion-financiera.php>), reportes, MVC, Single Page Application.

* Integración con APIs y web services de Bancard, Roshka.

Herramientas

* Backend: Python, Django Framework, Perl/CGI, Java, Java EE, JSP, PL/SQL, Oracle DB 11g, Spring Framework, Hibernate ORM.

* Frontend: JavaScript, JQuery, AngularJS, HTML, CSS, Bootstrap.

* Infraestructura: Docker, Docker-compose, Supervisor, uWSGI, Apache Server, Nginx, Tomcat, Wildfly; Windows, CentOS Linux, Red Hat Linux, Alpine Linux.

* Otros: Eclipse IDE, Tortoise SVN, Git, GitHub.

Primer período: 05/2013 a 09/2016

Tareas

* Desarrollo Full-Stack / Análisis de Requerimientos / Nuevos Productos / Innovación / Automatización de Procesos / Gestionar proveedores

de aplicaciones e infraestructura

Conocimientos adquiridos/aplicados

* Intranets, extranets, web services, servlets, API, middleware, banca web, banca móvil, medios de pago; SQL scripting, ERP "IT Gestión Financiera" (<http://www.bbsweb.com.ar/it-gestion-financiera.php>), reportes

Herramientas

* Perl/CGI, Perl, PHP, Java, Java EE, JSP, JavaScript, JQuery, ExtJS, HTML, CSS, Eclipse, Adobe, PL/SQL, Oracle 11g, Oracle forms and report 10g, JasperReports, Tortoise SVN, Apache, IIS, Tomcat, Jboss, Windows, CentOS Linux, Red Hat Linux

ANALISTA DE INNOVACION (2018 - 2019)

BASA - BANCO BASA

Carga Horaria: 40

Régimen: Dedicación total40

Otras Informaciones:

Analista de proyectos de innovación y desarrollo de TI Analista de proyectos de innovación y desarrollo de TI

TEXT MINING AND NATURAL LANGUAGE PROCESSING RESEARCH ENGINEER (2020 - 2021)

BCS - BARCELONA SUPERCOMPUTING CENTER

Carga Horaria: 40

Otras Informaciones:

Python, MongoDB, contenedores, Computación de alto rendimiento (HPC).

Técnico de Minería de Textos y PLN (Procesamiento de Lenguaje Natural) de Datos Sanitarios de la Unidad de Minería de Textos del Dpto. De Ciencias de la Vida (TeMU-BSC) financiado por el Plan de Impulso de las Tecnologías del Lenguaje de la Agenda Digital de la Secretaría de Estado de Telecomunicaciones e Información. (a) Adaptación e integración de componentes de PLN, (b) Desarrollo de mecanismos para la integración de recursos y datos en el contexto de aplicaciones de PLN, (c) Apoyo en el diseño de infraestructura de PLN biomédica / clínica, (d) Clasificación de textos / NER (Reconocimiento de Entidades Nombradas) de mensajes relacionados con la salud en las redes sociales (por ejemplo, Twitter), (e) Métodos de recuperación de información para obtener datos relevantes de las redes sociales. Aplicación de métodos automáticos para la recolección, extracción, representación, análisis y validación de Social Media Data para informática de la salud.

Actividades

Proyecto de Investigación y Desarrollo, BARCELONA SUPERCOMPUTING CENTER, Ministerio de Asuntos Económicos y Transformación Digital (1/2019 - 12/2012)

Plan de Impulso de las Tecnologías del Lenguaje (Sub-proyecto: Sanidad)

Participación: Integrante del Equipo

Descripción: Gobierno de España. Ministerio de Asuntos Económicos y Transformación Digital

Secretaría de Estado de Digitalización e Inteligencia Artificial

Secretaría de Estado de Telecomunicaciones e Infraestructuras Digitales

Plan de Impulso de las Tecnologías del Lenguaje: El Plan de Impulso de las Tecnologías del Lenguaje (Plan TL) tiene como objetivo fomentar el desarrollo del procesamiento del lenguaje natural, la traducción automática y los sistemas conversacionales en lengua española y lenguas cooficiales.

Sanidad

La Oficina Técnica de Sanidad del Plan TL trabaja principalmente en las siguientes líneas básicas de actuación:

Facilitar el acceso a recursos y herramientas de PLN que deben formar parte cualquier aplicación de PLN.

Garantizar la interoperabilidad necesaria entre los distintos recursos y herramientas y así garantizar la reutilización y mantenimiento de infraestructuras lingüísticas en el área de la biomedicina.

Establecer métodos de evaluación a través de campañas de evaluación comparativa de herramientas de PLN que potencien el desarrollo de infraestructuras lingüísticas en el área de la biomedicina de utilidad para el Plan.

Divulgar el proyecto, clave para el fomento y el desarrollo de las tecnologías del lenguaje en este ámbito.

Integrantes: Marvin M. Agüero-Torales(Responsable)

Situación: En Marcha; Tipo/Clase: Investigacion.

Alumnos: Pregrado (2); Especialización (1); Maestría Académica (4); Doctorado (8).

Financiadores: Ministerio de Asuntos Económicos y Transformación Digital - MINECO (Otra)

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / bioninformática
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / text mining
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / natural language processing
Humanidades / Lengua y Literatura / Lingüística / computational linguistics

Consultor Encargado de Informática (2013 - 2013)

BrSA - Britam S.A.

Carga Horaria: 48

Régimen: Dedicación total 48

Otras Informaciones:

Tareas

* Responsable de TI

Conocimientos adquiridos/aplicados

* Consultoría estratégica de infraestructura informática,
Migración de datos secuenciales de ERP Hansa World "Baltic" (<http://www.hansaworld.com>) a Microsoft SQL Server,
Extracción de conocimiento para toma de decisiones gerenciales,
Tareas administrativas

Herramientas

* Microsoft SQL Server, Mantis Bug Tracker; Servidores NAS, Samba, Proxy (FreeBSD Linux), Correo Electrónico (Zimbra), Thin Clients, Windows, CentOS Linux

FULL STACK / DEVOPS ENGINEER (2018 - 2018)

DECATHLON - DECATHLON

Carga Horaria: 20

Otras Informaciones:

VISA SPONSORSHIP AN RELOCATION PACKAGE COVERED BY DECATHLON.

Desarrollador Full Stack (2010 - 2011)

UH - Diario Última Hora

Carga Horaria: 48

Otras Informaciones:

Tareas

* Desarrollador Full-Stack / Análisis de Requerimientos / Nuevos Productos / Automatización de procesos

Conocimientos adquiridos/aplicados

* Sitios web, portales web, aplicaciones web, API, CMS: Wordpress, Joomla

Herramientas

* PHP, SQL, ASP, ASP .NET, HTML, JavaScript, JQuery, CSS, Perl, Microsoft SQL Server, MySQL, Apache, IIS, Windows, CentOS Linux, Adobe

Desarrollador Full Stack (2011 - 2011)

EDGE - EDGE S.A.

Carga Horaria: 45

Régimen: Dedicación total 45

Otras Informaciones:

Tareas

* Desarrollo Full-Stack / Análisis de Requerimientos / Innovación

Conocimientos adquiridos/aplicados

* Sistemas de información, aplicaciones web, aplicaciones de consola, extranets, soporte a usuarios, reportes

Herramientas

* C #, ASP .NET, SQL, HTML, JavaScript, CSS, PostgreSQL, PgSQL, IIS, Windows, Visual Studio, iTextSharp

Desarrollador Full Stack (2011 - 2012)

EBANSAE - Electroban S.A.E

Carga Horaria: 45

Régimen: Dedicación total45

Otras Informaciones:

Tareas

* Desarrollo Full-Stack / Análisis de Requerimientos / Innovación / Automatización de procesos

Conocimientos adquiridos/aplicados

* Aplicaciones de consola, aplicaciones web, web services, intranets, medios de pago; SQL scripting, reportes

Herramientas

* C#, SQL, Microsoft SQL Server, SyBase, ASP .NET, HTML, JavaScript, CSS, Ajax, Crystal Reports, LINQ, Entity Framework, ADO .NET, IIS, Windows, Visual Studio, Redmine

Desarrollador Full Stack (2009 - 2010)

EBANSAE - Electroban S.A.E

Carga Horaria: 48

Otras Informaciones:

Tareas

* Desarrollo Full-Stack (>Back-end) / Análisis de Requerimientos / Innovación / Automatización de procesos / Atención a Usuarios Helpdesk

Conocimientos adquiridos/aplicados

* Sistemas informáticos, aplicaciones web, intranets, ERP/BPM Bejerman (<http://www.bejerman.com/>), helpdesk, SQL Scripting, reportes

Herramientas

* C#, .NET, SQL, Microsoft SQL Server, ASP .NET, HTML, JavaScript, CSS, Visual Basic 6.0, Crystal Reports, Windows, Visual Studio

NLP & AI Applied Scientist (2022 - a la fecha)

Fujitsu - Fujitsu

Carga Horaria: 40

Régimen: Dedicación total40

Otras Informaciones:

Investigador Aplicado en el Centro de Excelencia (CoE) - Inteligencia de Datos de Fujitsu (tercera en el ranking mundial de servicios TIC). Trabajando en el proyecto de medicina personalizada para estudios genómicos de un individuo: "Medigenomics", de la Consejería de Sanidad de la Comunidad de Madrid y el Ministerio de Ciencia e Innovación de España (Fondo Europeo de Desarrollo Regional-FEDER). Minería de texto y PNL sobre historias clínicas electrónicas/digitales y literatura médica sobre información genómica: voz-a-texto, Reconocimiento óptico de caracteres (OCR), reconocimiento de entidades nombradas (NER), clasificación de documentos, etc.

Applied Scientist at the Centre of Excellence (CoE) - Data Intelligence, GenAI & NLP. I am part of Fujitsu's AIDP team. Working on the LEIA project of the Royal Spanish Academy (RAE), funded by the PERTE - New economics of language of the Secretary of State of Telecommunications and the Information

Consultor SAP Portal (2012 - 2013)

INFOCENTER - INFOCENTER

Carga Horaria: 45

Régimen: Dedicación total45

Otras Informaciones:

Tareas

* Consultor Técnico SAP Portal Java / Consultor SAP Basis / Gestión de proveedores de aplicaciones

Conocimientos adquiridos/aplicados

* ERP SAP NetWeaver Portal (<http://www.sap.com>), SAP Web Dynpro Java, Eclipse, SAP HCM: SAP ESS y SAP MSS, SAP SRM, SAP BASIS básico

Herramientas

* Java, J2EE, SQL, Microsoft SQL Server, MaxDB, Windows, Linux

Desarrollador de back-end/front-end (2008 - 2009)

SDAPYSA - S.D.A. Paraguay S.A.

Carga Horaria: 40

Otras Informaciones:

Tareas

* Desarrollo back-end y front-end / Análisis de Requerimientos / Helpdesk

Conocimientos adquiridos/aplicados

* Sistemas de información, aplicaciones web, intranets, helpdesk, reportes

Herramientas

* Visual Basic 6.0, .NET, SQL, Pg/SQL, PHP, HTML, JavaScript, CSS, Microsoft SQL Server, MySQL, PostgreSQL, Microsoft Office Access, helpdesk, Windows, Visual Studio

Investigador asociado (2024 - a la fecha)

UAA - Universidad Autónoma de Asunción

Carga Horaria: 3

Otras Informaciones:

Participación en el proyecto: Optimización de modelos de Aprendizaje Profundo para clasificación de imágenes y texto.

(PINV01-401)

Co-tutor de tesis de maestría (2023 - a la fecha)

UCOM - Universidad Comunera

Carga Horaria: 2

Otras Informaciones:

Junto a Margarita Ruiz-Olazar, sobre estudio de factibilidad de grandes modelos de lenguajes para el guaraní/jopará

Investigador invitado (2020 - 2021)

UCA - Universidad de Cádiz

Carga Horaria: 38

Otras Informaciones:

Ingeniero de PLN (Procesamiento de lenguaje natural), científico de datos, Python stack.

Proyecto: Deep Learning y Procesamiento de Lenguaje Natural para el tratamiento automatizado de informes clínicos (financiado por Instituto de Salud Carlos III).

Actividades

Proyecto de Investigación y Desarrollo, Universidad de Cádiz, Instituto de Salud Carlos III, Majadahonda, Madrid (12/2020 - 1/2021)

Automatic processing of clinical reports applying deep learning and NLP

Participación: Integrante del Equipo

Descripción: There is a large amount of extremely valuable information stored in clinical notes written in plain unstructured text. This project aims to develop NLP tools for automated treatment of clinical reports, structuring the text into medical terminologies such as ICD-10 or SNOMED-CT. UCA Datalab collaborates with the Hospital Puerta del Mar in Cádiz, where Dr. García Rojo, Head of Pathology, provides us with more than 160K clinical notes on biopsies, anotated in SNOMED, in order to train NLP algorithms. The possible outcomes of this project are very rich and include efficient cohort segmentation for clinical trials, information retrieval and faster, reliable coding of structured reports during clinical visits.

Integrantes: Marvin M. Agüero-Torales(Responsable)

Situación: En Marcha; Tipo/Clase: Investigación.

Alumnos: Pregrado (1); Maestría Académica (2); Doctorado (3).

Financiadores: Instituto de Salud Carlos III, Majadahonda, Madrid - iscii (Apoyo financiero)

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSING
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / BIOINFORMATIC
Ciencias Naturales / Matemáticas / Matemática Aplicada / MACHINE LEARNING

Colaborador NLP (2022 - a la fecha)

UG - Universidad de Granada

Carga Horaria: 2

Otras Informaciones:

Colaborador en Andalusian Research Institute in Data Science and Computational Intelligence dasci.es y el grupo <https://sci2s.ugr.es/>*Actividades*

Líneas de Investigación (2/2022 - a la fecha)

NATURAL LANGUAGE PROCESSING

Participación: Integrante del Equipo

Descripción: In this line, we take the direction of text mining and natural language processing, at the intersection between computation, artificial intelligence, and computational linguistics, with a clear emphasis on multilingualism in low-resource languages. The objective is to provide a better understanding of this problem (or related approaches) as well as a holistic analysis and view of it.

1. Study the different existing machine-learning approaches, that aim at the language-independence, especially the most current ones based on neural networks, with respect to multilingual opinions written in social media, even if were written with code-switching.
2. Develop new linguistic resources for text analysis in low-resource languages and dialects, especially written in social media.
3. Build machine learning models for NLP of low-resource languages and dialects, especially in monolingual, multilingual, and code-switching settings.

Palabras Clave: nlp; text mining; low-recource languages; multilinguality; code-switching; natural language understanding; twitter; tripadvisor; noticias; social media;

Integrantes: Marvin M. Agüero-Torales(Responsable)

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSING
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / ARTIFICIAL INTELLIGENCE
Humanidades / Lengua y Literatura / Lingüística / COMPUTATIONAL LINGUISTICS

Doctorando (2017 - 2022)

UG - Universidad de Granada

Carga Horaria: 20

Otras Informaciones:

Python, R, NoSQL, Contenedores.

Captura, Preprocesamiento, Análisis y Visualización de Datos Web y Redes Sociales como Twitter o TripAdvisor, así como proceso de anotación de datos: (a) Desarrollo de una herramienta en la Nube para Análisis de Sentimiento de textos redactados en inglés en los dominios de turismo (gastronomía), (b) Revisión del estado del arte del Análisis de Sentimiento Multilingüe sobre opiniones en Redes Sociales con Deep Learning, (c) Aplicar Modelado de Temas para descubrir temas en Twitter sobre el brote de COVID-19 en España, (d) Proponer análisis de sentimientos (y tareas relacionadas a nivel de oración y palabra) utilizando técnicas de aprendizaje automático en opiniones de redes sociales escritas en idiomas multilingües, code-switching o de bajos recursos (guaraní/jopará). (e) Creación de modelos del estado del arte para el procesamiento del idioma guaraní: transformer - BERT (<https://huggingface.co/mmagüero>)

TECNICO DE APOYO A LA INVESTIGACION (2019 - 2019)

UdC - Universidad de la Coruña

Carga Horaria: 38

Régimen: Dedicación total38

Otras Informaciones:

Python, R, PostgreSQL, Tableau, contenedores.

(a) Control predictivo de acuerdos de nivel de servicio (SLAs), (b) Detección de errores en la captura de datos de campo, (c) Recomendaciones al trabajo de campo de los técnicos, (d) Clasificación automática de texto en Lenguaje Natural. Exploración e implementación de técnicas de Machine Learning y Data Mining que permitan el desarrollo de funcionalidades de asistencia inteligente y que finalmente fueron integradas y validadas dentro de un sistema real en uso por una multinacional de servicios de autos expertos (DEKRA: STEPS financiado por el Programa Estatal de I+D+i Orientada a los Retos de la Sociedad).

Actividades

Proyecto de Investigación y Desarrollo, Universidad de la Coruña, AGENCIA ESTATAL DE INVESTIGACIÓN (6/2019 - 12/2019)

STEPS: Solucións Tecnolóxicas para a Evolución na Prestación de Servizos en campo

Participación: Integrante del Equipo

Descripción: Exploración e implementación de técnicas de Machine Learning y Data Mining que permitan el desarrollo de funcionalidades de asistencia inteligente y que finalmente fueron integradas y validadas dentro de un sistema real en uso por una multinacional de servicios de autos expertos (DEKRA: STEPS financiado por el Programa Estatal de I+D+i Orientada a los Retos de la Sociedad).

Integrantes: Marvin M. Agüero-Torales(Responsable)

Situación: Concluido; Tipo/Clase: Investigacion.

Alumnos: Pregrado (2); Maestría Académica (3); Maestría profesionalizante (2); Doctorado (6).

Financiadores: AGENCIA ESTATAL DE INVESTIGACIÓN - AEI (Apoyo financiero)

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / MACHINE LEARNING
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / DATA MINING
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSING
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / COMPUTER VISION

RESEARCH ENGINEER (2021 - 2022)

UOC - Universitat Oberta de Catalunya

Carga Horaria: 38

Régimen: Dedicación total38

Otras Informaciones:

Ingeniero de PLN, Data Wrangler: Python stack, PostgreSQL, Web Scraping.

Ingeniero Técnico de "Redes Sociales del Pasado. Mapeo de la modernidad literaria hispana y lusófona, 1898-1959" (ERC ID 803860).

Participo en el diseño de bases de datos; validar, limpiar, estandarizar y completar los datos recopilados; anotar datos; organizar bases de datos documentales. Minería de texto y PLN: recuperación de texto, modelado de temas, NER, tokenización, etc.

Actividades

Proyecto de Investigación y Desarrollo, Universitat Oberta de Catalunya, Unión Europea (2/2021 - a la fecha)

Social Networks of the Past: Mapping Hispanic and Lusophone Literary Modernity, 1898-1959

Participación: Integrante del Equipo

Descripción: Starting Grants | ERC: European Research Council - What do we know about European and non-European social and cultural life of the 20th century? What was the impact of Hispanic and lusophone literary networks and cultural mediators on international modernity between 1898 and 1959? The EU-funded MapModern project will use cutting-edge technology to answer these questions. It will reveal the lost history of Ibero-American mediators in modernist intercultural and multilingual networks. The project's twofold methodology includes, first, an open and collaborative research tool providing a data source for quantitative and qualitative analysis on Ibero-American mediators. The second focuses on the cultural transformation process, exploring the institutionalisation of Ibero-American cultures and the rise of translated literature in key Ibero-American modernist journals. The position of Ibero-American women in the cultural field will also be studied.
<https://cordis.europa.eu/project/id/803860/es>

Integrantes: Marvin M. Agüero-Torales(Responsable)

Situación: En Marcha; Tipo/Clase: Investigacion.

Alumnos: Pregrado (1); Maestría Académica (4); Doctorado (8).

Financiadores: Unión Europea - UE (Beca)

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / DATA MINING
Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / BIG DATA
Humanidades / Otras Humanidades / DIGITAL HUMANITIES

Significado de su trabajo en el contexto de los principales problemas planteados en su área:

Mi línea de investigación se centra principalmente en el área de Procesamiento de Lenguaje Natural, en la intersección entre la computación, la inteligencia artificial y la lingüística computacional. Más específicamente en la aplicación de técnicas de machine learning para procesar textos escritos multi-idiomas, particularmente en lenguajes con escasez de recursos computacionales (como el Guaraní) así como los que sufren de code-switching (alternancia de una o más lenguas en un texto, incluso a nivel de palabra, por ej. el Jopará).

Dentro de esto, mi trabajo se enfoca precisamente en la investigación, creación y desarrollo de modelos de machine learning (especialmente los considerados como del estado arte: deep learning del tipo recurrentes o transformer, entre otros), enfoques, técnicas y recursos tecnológicos (como corpus lingüísticos, datasets, léxicos o herramientas para procesar este tipo de idiomas).

En cuanto a las tareas de Procesamiento de Lenguaje Natural, he versado mi trabajo sobre el análisis de sentimientos y detección de afectos, conocidas como computación afectiva, a saber: clasificación de la polaridad del texto, detección de humor, reconocimiento de emociones, identificación de lenguaje ofensivo, tóxico, soez o racista, entre otros. Útil para estudios en medios sociales, como pueden ser la polarización de campañas (por ej., marketing o políticas). Además, del NER - reconocimiento de entidades nombradas dentro de un texto, especialmente las que conciernen al ámbito de la salud y la biomedicina, como pueden ser reconocer enfermedades, síntomas, especies, etc, incluso aplicados a finanzas, en un texto dado; así como más generales para el guaraní (idioma, lugar, persona, organización). A modo

general, me he centrado, en la clasificación de textos a nivel de palabra y sentencia, su agrupación y modelado de temas, así como la creación de incrustaciones (embeddings) de objetos textuales (documentos, palabras, sub-palabras, caracteres), sean o no contextualizadas, o la creación y ajuste de grandes modelos de lenguaje para casos de uso específicos (como salud o biomedicina) o idiomas concretos (como el guaraní o el jopará). Siempre enmarcado en el estudio de idiomas de escasos recursos y que presentan multilingüismo, aunque también en el idioma español e inglés.

Producción Técnica

Cursos de corta duración dictados

1 **Marvin M. Agüero-Torales; Sebastian Grillo; Diploma in Artificial Intelligence Applied to Products and Services (Certificate, Polytechnic Faculty - National University of Asunción - FPUNA), 2024. (Otro)**

Palabras Clave: nlp; machine learning; python; data management;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

Referencias adicionales: Paraguay/Español; Medio: Internet.

Institución promotora/financiadora: Universidad Nacional de Asunción - Facultad Politécnica

Participación: Docente. Duración: 5 meses.

Observaciones: Módulos:

Procesamiento de Lenguaje Natural

2 **Marvin M. Agüero-Torales; Sebastian Grillo; Machine Learning Diploma (Certificate, Ministry of Information Technologies and Communications - MITIC & FPUNA), 2023. (Especialización)**

Palabras Clave: unsupervised learning; final project; data management; supervised learning; machine learning;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA

Referencias adicionales: Paraguay/Español; Medio: Internet.

Institución promotora/financiadora: Universidad Nacional de Asunción - Facultad Politécnica

Participación: Docente. Duración: 6 meses.

Observaciones: Profesor de los módulos de:

unsupervised learning

final project

3 **Marvin M. Agüero-Torales Taller en Data Science, 2018. (Perfeccionamiento)**

Palabras Clave: sentiment analysis; web scraping; social media; text mining; natural language processing; data science; data mining;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática /

Referencias adicionales: Paraguay/Español; Medio: Otros.

Institución promotora/financiadora: Universidad Comunera

Participación: Docente. Lugar: Universidad Comunera. Ciudad: Asunción.

Sitio Web: www.ucom.edu.py/eventos/taller-de-data-science/

Observaciones: Aplicar el ciclo completo de la ciencia de datos en un caso práctico real.

Adquirir habilidades en el manejo de herramientas computacionales para facilitar los descubrimientos al organizar, visualizar y explorar datos.

Equipo de instructores

M.Eng. Marvin Agüero

Dr. Daniel Romero

D.Sc. Margarita Ruiz

Obtener, organizar, visualizar y explorar datos usando Python

Algoritmos de data mining

Aplicación de un caso práctico real para el análisis de datos masivos procedentes de medios sociales.

Procesos o técnicas

1 **Marvin M. Agüero-Torales; DAVID VILARES; A.G. López-Herrera; Jopara (Guarani-dominant mixed with Spanish) text-based affect detection corpora , 2022.**

Referencias adicionales: España;/ Medio: Internet.

Finalidad: Entrenar modelos para detección de afecto en textos cortos. Disponibilidad: irrestricta.

Institución promotora/financiadora: Universidad de Granada.

Sitio Web: to be appear

Observaciones: Dimensiones de afecto cubiertas:

Emociones.

Humor

Lenguaje tóxico/ofensivo y discurso de odio

2 Alejandro Asensio; Antonio Miranda-Escalada; Marvin M. Agüero-Torales; Martin Krallinger; Occupations gazetteer - ProfNER & MEDDOPROF - occupations, professions and working status terms with their associated codes, 2021.

Referencias adicionales: España/; Medio: Internet.

Finalidad: Diccionario de profesiones para detectarlos luego en un texto por coincidencia. Disponibilidad: irrestricta.

Institución promotora/financiadora: BARCELONA SUPERCOMPUTING CENTER.

Sitio Web: <https://doi.org/10.5281/zenodo.4524658>

Observaciones: Gazetteer of occupations extracted from a set of terminologies (DeCS, ESCO, SnomedCT and WordNet). In addition, we include in the gazetteer a list of occupations detected by Stanford CoreNLP in a large collection of social media profiles. For every term, we include its lemma. The lemma was extracted using the PlanTL Med_Tagger: https://github.com/PlanTL-SANIDAD/SPACCC_POS-TAGGER

3 Marvin M. Agüero-Torales; DAVID VILARES; A.G. López-Herrera; Jopara (Guarani-dominant mixed with Spanish) sentiment analysis corpus , 2021.

Referencias adicionales: España/; Medio: Internet.

Finalidad: Corpus a utilizar para detección de sentimiento en textos cortos. Disponibilidad: irrestricta.

Institución promotora/financiadora: Universidad de Granada.

Sitio Web: <https://doi.org/10.7910/DVN/GLDX14>

Observaciones: Corpus for the work: On the logistical difficulties and findings of Jopara Sentiment Analysis. In Proceedings on CALCS 2021 (co-located with NAACL 2021) - Fifth Workshop on Computational Approaches to Linguistic Code Switching

4 Antonio Miranda-Escalada; Vicent Briva-Iglesias; Eulàlia Farré-Maduell; Salvador Lima-López; Marvin M. Agüero-Torales; Martin Krallinger; ProfNER corpus: gold standard annotations for profession detection in Spanish COVID-19 tweets, 2020.

Referencias adicionales: España/; Medio: Internet.

Finalidad: Corpus para entrenar modelos de machine learning capaces de reconocer profesiones en el texto. Disponibilidad: irrestricta.

Institución promotora/financiadora: BARCELONA SUPERCOMPUTING CENTER.

Sitio Web: <https://doi.org/10.5281/zenodo.4309356>

Observaciones: Gold Standard annotations for SMM4H-Spanish shared task and unannotated test and background files. SMM4H 2021 accepted at NAACL (scheduled in Mexico City in June) <https://2021.naacl.org/>.

5 Marvin M. Agüero-Torales; DAVID VILARES; A.G. López-Herrera; COVID-19-Outbreak-Tweets-Spain-es (COTwes), 2020.

Referencias adicionales: España/; Medio: Internet.

Finalidad: Corpus de tweets sobre covid-19 sin etiquetas para métodos no supervisados. Disponibilidad: irrestricta.

Institución promotora/financiadora: Universidad de Granada.

Sitio Web: <https://doi.org/10.7910/DVN/6PPSAZ>

Observaciones: Trabajo relacionados con: Discovering topics in Twitter about the COVID-19 outbreak in Spain. Procesamiento Del Lenguaje Natural, 66, 177-190. Recuperado de <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6333>

Productos tecnológicos

1 Marvin M. Agüero-Torales; DAVID VILARES; Guarani BERT Representations, 2022.

Referencias adicionales: España/; Medio: Internet.

Finalidad: Precedir palabras en guaraní/jopará según el contexto, útiles para fine-tuning de otras tareas.

Institución promotora/financiadora: Universidad de Granada.

Sitio Web: <https://huggingface.co/mmaguero>

Observaciones: Modelo que ayuda a los ordenadores a entender el significado de un lenguaje ambiguo en el texto, utilizando el texto circundante para establecer el contexto. Basados en transformers del tipo BERT <https://aclanthology.org/N19-1423/>

2 Marvin M. Agüero-Torales Tweets downloader, 2021.

Palabras Clave: web scraping; api; twitter; text analysis;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática /

Referencias adicionales: España/; Medio: Internet.

Finalidad: Descargar tweets vía la API o través de web scraping. Disponibilidad: irrestricta.

Institución promotora/financiadora: Universidad de Granada.

Sitio Web: <https://github.com/mmaguero/tweets-downloader>

Observaciones: Download tweets by Twitter Official API or SNScrape

Input: a txt with terms/words to search (as keywords).

3 Antonio Miranda-Escalada; Marvin M. Agüero-Torales; Martin Krallinger; Spanish COVID-19 Twitter Embeddings in FastText, 2021.

Referencias adicionales: España/; Medio: Internet.

Finalidad: Usar con modelos de deep learning para clasificar textos en español sobre covid-19. Disponibilidad: irrestricta.

Institución promotora/financiadora: BARCELONA SUPERCOMPUTING CENTER.

Sitio Web: <https://doi.org/10.5281/zenodo.4449929>

Observaciones: Intro 300-dimensional FastText embeddings generated from 140 million tweets in Spanish. All tweets are COVID19-related, meaning that they include one or more keywords related to COVID-19 and lockdown.

4 Marvin M. Agüero-Torales; Jorge Saldivar; Language detection (including Guarani) , 2020.

Palabras Clave: language detection; guarani;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSING / Humanidades / Lengua y Literatura / Lingüística / COMPUTATIONAL LINGUISTICS

Referencias adicionales: España/; Medio: Internet.

Finalidad: Language detection (majority voting), including Guarani language.. Disponibilidad: irrestricta.

Institución promotora/financiadora: BARCELONA SUPERCOMPUTING CENTER.

Sitio Web: https://github.com/mmaguero/lang_detection

Observaciones: Python based software:

Use: : given a text, the software determines the language of the text.

Tools used for this purpose...

```
polyglot*,
fastText*,
langdetect,
langid, and
textcat*.
```

*: Supports the Guarani language.

5 Marvin M. Agüero-Torales Lexico afinn personalizado al español, 2019.

Palabras Clave: sentiment analysis; afinn; lexico; diccionario; lenguaje natural; text mining;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Ciencias Sociales / Comunicación y Medios / Ciencias de la Información /

Referencias adicionales: España/; Medio: Internet.

Finalidad: Analizar sentimientos en un texto dado, polaridad de -1 a 1, siendo 1 positivo, -1 negativo.. Disponibilidad: irrestricta.

Institución promotora/financiadora: Universidad de la Coruña.

Sitio Web: <https://github.com/mmaguero/afinn>

Observaciones: AFINN sentiment analysis in Python: Wordlist-based approach for sentiment analysis.

6 Marvin M. Agüero-Torales Gastro-miner, 2017.

Palabras Clave: análisis de sentimientos; redes sociales; tripadvisor; minería de opiniones;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Lenguaje Natural, Minería de Opiniones

Referencias adicionales: España/; Medio: Internet.

Finalidad: Análisis de sentimientos de opiniones del rubro gastronómico. Disponibilidad: irrestricta.

Institución promotora/financiadora: Universidad de Granada.

Sitio Web: <https://github.com/mmaguero/cloud-based-tool-SA>

Observaciones: A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor
doi.org/10.1016/j.procs.2019.12.002

7 Marvin M. Agüero-Torales API de Medios de Pago, 2016.

Palabras Clave: api; medios de pago; banca online; pago electrónico;

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Middleware/API Ciencias Sociales / Economía y Negocios / Negocios y Administración / Pago Electrónico

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: API de pagos electrónicos de servicios. Disponibilidad: restricta.

Institución promotora/financiadora: Banco Atlas S.A..

Sitio Web: <https://secure.atlas.com.py/inicio.html>

Observaciones: *API embebida en banca online como un middleware de a su vez la API de pagos de Roshka S.A.:

-Funciona como un API de APIs, un middleware entre la API de Roshka S.A. y la banca online, la banca móvil y el sistema de gestión.

-Convierte formularios abstractos de WS SOAP a XML y viceversa, desde Oracle 11g

-Los usuarios pueden pagar cualquier servicio, público o privado.

8 Marvin M. Agüero-Torales Extranet Banco Atlas: Solicitud de créditos, 2015.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Extranet Ciencias Naturales / Matemáticas / Matemática Aplicada /

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: Lograr mayor eficiencia en la entrega de las tarjetas de créditos, automatizando el proceso, reduciendo los costos actuales con mayor seguridad y tiempo.. Disponibilidad: restringida.

Institución promotora/financiadora: Banco Atlas S.A..

Sitio Web: <https://secure.atlas.com.py/gfweb.html>

Observaciones: *Web responsiva: adaptable a tamaños y resoluciones de pantalla

*Pantalla de alta de solicitudes crediticias, datos del cliente

*Obtención automática de informaciones crediticias de clientes (mediante la implementación de servicios web de Informconf S.A.)

*Scoring crediticio automático (mediante la implementación de servicios web de Datalab Consulting S.A.).

*Embozado instantáneo de tarjetas de créditos, utilizando la impresora NBS Technologies Advantage M20, habilitándola automáticamente para uso (mediante servicios web de Bancard S.A.).

9 Marvin M. Agüero-Torales; M. Trinidad; Aplicación de Cifrado de archivos para lotes de pagos electrónicos, 2014.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Extranet

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: Cifrado de archivos para lotes de pagos electrónicos de empresas a proveedores y salarios del Home Banking. Disponibilidad: restringida.

Institución promotora/financiadora: Banco Atlas S.A..

Sitio Web: <https://secure.atlas.com.py/inicio.html>

Observaciones: Cifrado GPG de archivos para lotes de pagos electrónicos de empresas a proveedores y salarios desde el Home Banking del Banco Atlas S.A., mediante Perl, GnuPG y Linux

10 Marvin M. Agüero-Torales Sistema de Requerimientos Informáticos, 2013.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Intranet

Referencias adicionales: Paraguay/; Medio: Otros.

Finalidad: Sistema de Requerimientos del área de Tecnología, seguimiento y control de requerimientos y desarrolladores. Disponibilidad: restringida.

Institución promotora/financiadora: Banco Atlas S.A..

Sitio Web: <http://www.bancoatlas.com.py>

Observaciones: Sistema de Requerimientos del área de Tecnología de la entidad:

*Alta de requerimientos

*Control y gestión de requerimientos y personas

*Diagrama Gantt

*Interacción de usuarios solicitantes y responsables del desarrollo de requerimientos

11 Marvin M. Agüero-Torales; N. Negraes; A. Tello; Sistema de Análisis de Riesgos Crediticios, 2012.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Intranet Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Cálculo de scoring crediticio, reglas crediticias

Referencias adicionales: Paraguay/; Medio: Otros.

Finalidad: Análisis y otorgamientos de créditos a bancas minoristas, micropymes y personas. Disponibilidad: restringida.

Institución promotora/financiadora: Electroban S.A.E.

Sitio Web: <http://electroban.com.py/>

Observaciones: Sistema Web (Intranet) de Análisis de Riesgos Crediticios de la Empresa Electroban S.A.E.:

*Aplicación e implementación del API de Google Maps para localizaciones de clientes, personales y sucursales de la empresa.

*Módulo inteligente de "Reglas de Créditos" parametrizable.

*Scoring crediticio propio.

*Módulo de "Cobranzas Express" online (mediante servicio proveído por The@Derby [Aliado Móvil]).

*Obtención automática de informaciones crediticias de clientes (mediante la implementación del servicio "Host To Host" de Informconf S.A.)

*Módulo de "Datos Históricos del Cliente" (datos laborales, direcciones, teléfonos, operaciones crediticias, etc.).

12 Marvin M. Agüero-Torales Web Service para pagos online de cuotas de clientes de empresa comercial, 2012.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Web Services

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: Web Service para pagos online de cuotas de clientes de la empresa Electroban S.A.E. . Disponibilidad: restringida.

Institución promotora/financiadora: Electroban S.A.E.

Sitio Web: <http://www.electroban.com.py>

Observaciones: Desarrollo del Web Service para pagos online de cuotas de clientes de la empresa Electroban S.A.E. (mediante bocas de cobranzas provistas por Pronet S.A. [Aquí Pago] y Documenta S.A. [Practipago])

13 Marvin M. Agüero-Torales; A. Acosta; Consultas Médicas, 2011.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Web 2.0, Extranet

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: "Consultas Médicas" telefónicas de TIGO. Disponibilidad: restringida.

Institución promotora/financiadora: EDGE S.A..

Sitio Web: <http://salud.edge.com.py>

Observaciones: Desarrollo del Sistema Web "Consultas Médicas" telefónicas de TIGO, facturación de llamadas hechas por pacientes a una línea exclusiva del servicio. Emisión de comprobantes y extractos de línea.

14 Marvin M. Agüero-Torales Blogósfera Paraguay, 2010.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Web 2.0, CMS

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: Un directorio de blogs nacionales. Disponibilidad: irrestricta.

Institución promotora/financiadora: Diario Última Hora.

Sitio Web: blogosfera.ultimahora.com

Observaciones: Un directorio de blogs nacionales, herramienta para el concurso de "Los 10 mejores blogs del Paraguay" realizado del año 2010 al 2011

15 Marvin M. Agüero-Torales; Z. Sanabria; Telefuturo "Somos", 2010.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Web 2.0, CMS

Referencias adicionales: Paraguay/; Medio: Internet.

Finalidad: Sitio web institucional de la teledifusora Telefuturo. Disponibilidad: irrestricta.

Institución promotora/financiadora: Telefuturo.

Sitio Web: <http://www.telefuturo.com.py>

Observaciones: Sitio web del año 2010, campaña publicitaria TELEFUTURO "Somos"

16 Marvin M. Agüero-Torales; N. Negraes; Sistema de Liquidación de Haberes y Gestión del Personal , 2009.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Intranet

Referencias adicionales: Paraguay/; Medio: Otros.

Finalidad: Liquidación de Haberes y Gestión del Personal . Disponibilidad: restringida.

Institución promotora/financiadora: Electroban S.A.E.

Sitio Web: <http://www.electroban.com.py>

Observaciones: Desarrollo del Sistema Web (Intranet) de Liquidación de Haberes y Gestión del Personal de la Empresa Electroban S.A.E., cálculo de comisiones, salarios y honorarios, vacaciones, etc.

17 Marvin M. Agüero-Torales Sistema de Facturación y Control de Stock, 2008.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Intranet

Referencias adicionales: Paraguay/; Medio: Otros.

Finalidad: Facturación y Control de Stock para empresas del área de gastronomía. Disponibilidad: restringida.

Institución promotora/financiadora: S.D.A. Paraguay S.A..

Observaciones: Desarrollo de un Sistema Web de Facturación y Control de Stock para empresas del área de gastronomía, punto de venta, caja, facturación, stock, comandas, propinas, etc.

18 Marvin M. Agüero-Torales; R. Agüero; Sistema de Liquidación de Haberes, 2008.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Intranet

Referencias adicionales: Paraguay/; Medio: Otros.

Finalidad: Liquidación de Haberes para Entes Gubernamentales. Disponibilidad: restringida.

Institución promotora/financiadora: S.D.A. Paraguay S.A..

Observaciones: Desarrollo de Sistema Web de Liquidación de Haberes para Entes Gubernamentales, asalariados, contratados, jornaleros.

Producción Bibliográfica

Trabajos en eventos

Trabajos completos en anales de eventos

- 1 **Ona De Gibert; Robert Pugh; Ali Marashian; Raul Vazquez; Abteen Ebrahimi; PAVEL DENISOV; Enora Rice; Edward Gow-Smith; Juan Prieto; Melissa Robles; Rubén Manrique; Oscar Moreno; Angel Lino; Rolando Coto-Solano; Aldo Alvarez; Marvin M. Agüero-Torales; John E. Ortega; Luis Chiruzzo; Arturo Oncevay; Shruti Rijhwani; Katharina von der Wense; MANUEL MAGER**, **Metrics of the AmericasNLP 2025 Shared Tasks on Machine Translation, Creation of Educational Material, and Translation Metrics for Indigenous Languages of the Americas**. In: **Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP), 2025 Proceedings of the Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP). 2025.**

Medio: Internet.

Sitio Web: <https://aclanthology.org/2025.americasnlp-1.16/>

Observaciones: Ona De Gibert, Robert Pugh, Ali Marashian, Raul Vazquez, Abteen Ebrahimi, Pavel Denisov, Enora Rice, Edward Gow-Smith, Juan Prieto, Melissa Robles, Rubén Manrique, Oscar Moreno, Angel Lino, Rolando Coto-Solano, Aldo Alvarez, Marvin Agüero-Torales, John E. Ortega, Luis Chiruzzo, Arturo Oncevay, Shruti Rijhwani, Katharina Von Der Wense, and Manuel Mager. 2025. (<https://aclanthology.org/2025.americasnlp-1.16/>)

- 2 **Agustín Lucas; Alexis Baladón; Victoria Pardiñas; Marvin M. Agüero-Torales; Santiago Góngora; Luis Chiruzzo; Grammar-based Data Augmentation for Low-Resource Languages: The Case of Guarani-Spanish Neural Machine Translation**. In: **2024 Conference of the NAACL: Human Language Technologies, 2024 Mexico City, Mexico Proceedings of the 2024 Conference of the NAACL: Human Language Technologies (Volume 1: Long Papers). 2024.**

Medio: Internet.

Sitio Web: <https://aclanthology.org/2024.naacl-long.354/>

Observaciones: Agustín Lucas, Alexis Baladón, Victoria Pardiñas, Marvin Agüero-Torales, Santiago Góngora, and Luis Chiruzzo. 2024. Grammar-based Data Augmentation for Low-Resource Languages: The Case of Guarani-Spanish Neural Machine Translation. In Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers), pages 6385-6397, Mexico City, Mexico. Association for Computational Linguistics.

- 3 **Luis Chiruzzo; PAVEL DENISOV; ALEJANDRO MOLINAS-VILLEGAS; SILVIA FERNANDEZ-SABIDO; Rolando Coto-Solano; Marvin M. Agüero-Torales; Aldo Alvarez; Samuel Canul-Yah; Lorena Hau-Ucán; Abteen Ebrahimi; Robert Pugh; Arturo Oncevay; Shruti Rijhwani; Katharina von der Wense; MANUEL MAGER**, **Findings of the AmericasNLP 2024 Shared Task on the Creation of Educational Materials for Indigenous Languages**. In: **4th Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP 2024), 2024 Proceedings of the 4th Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP 2024). 2024.**

Medio: Internet.

Sitio Web: <https://aclanthology.org/2024.americasnlp-1.27/>

Observaciones: Luis Chiruzzo, Pavel Denisov, Alejandro Molina-Villegas, Silvia Fernandez-Sabido, Rolando Coto-Solano, Marvin Agüero-Torales, Aldo Alvarez, Samuel Canul-Yah, Lorena Hau-Ucán, Abteen Ebrahimi, Robert Pugh, Arturo Oncevay, Shruti Rijhwani, Katharina von der Wense, and Manuel Mager. 2024.

[Findings of the AmericasNLP 2024 Shared Task on the Creation of Educational Materials for Indigenous Languages](<https://aclanthology.org/2024.americasnlp-1.27/>) (Chiruzzo et al., AmericasNLP 2024)

- 4 **Luis Chiruzzo; Marvin M. Agüero-Torales; Aldo Alvarez; Yliana Rodriguez; Initial Experiments for Building a Guarani WordNet**. In: **GWC 2023, 2023 San Sebastian, Donostia, País Vasco, España Proceedings of 12th International Global Wordnet Conference (GWC2023). 2023.**

Medio: Otros.

Sitio Web: <https://www.hitx.eus/gwc2023/programme>

- 5 **Luis Chiruzzo; Santiago Góngora; Aldo Alvarez; Gustavo Giménez-Lugo; Marvin M. Agüero-Torales; Yliana Rodriguez; Jojajovai: A Parallel Guarani-Spanish Corpus for MT Benchmarking**. In: **LREC, 2022 Marsella, Francia Proceedings of the Thirteenth Language Resources and Evaluation Conference. 2022.**

Medio: Otros.

Sitio Web: <https://aclanthology.org/2022.lrec-1.226/>

- 6 **Marvin M. Agüero-Torales; DAVID VILARES; A.G. López-Herrera; On the logistical difficulties and findings of Jopara Sentiment Analysis**. In: **Fifth Workshop on Computational Approaches to Linguistic Code-Switching (co-located in NAACL), 2021 Mexico (Online) Proceedings of the Fifth Workshop on Computational Approaches to Linguistic Code-Switching. 2021.**

Medio: Internet.

Sitio Web: <https://aclanthology.org/2021.calcs-1.12/>

Observaciones: DOI: 10.18653/v1/2021.calcs-1.12

Code: <https://github.com/mmaguero/josa-corpus>

- 7 **Antonio Miranda-Escalada; Eulàlia Farré-Maduell; Salvador Lima-López; Luis Gascó ; Vicent Briva-Iglesias; Marvin M. Agüero-Torales; Martin Krallinger; The ProfNER shared task on automatic recognition of occupation mentions in social media: systems, evaluation, guidelines, embeddings and corpora**. In: **Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task (co-located in NAACL), 2021 Mexico (Online) Proceedings of the Sixth Social Media Mining for Health (#SMM4H) Workshop and Shared Task. 2021.**

Medio: Internet.

Sitio Web: <https://aclanthology.org/2021.smm4h-1.3/>

Observaciones: To evaluate current NLP methods and to generate resources, we have organized the ProfNER track at SMM4H 2021, providing ProfNER participants with a Gold Standard corpus of manually annotated tweets (human IAA of 0.919) following annotation guidelines available in Spanish and English, an occupation gazetteer, a machine-translated version of tweets, and FastText embeddings. Out of 35 registered teams, 11 submitted a total of 27 runs. Corpus: <https://doi.org/10.5281/zenodo.4309356>

8 Marvin M. Agüero-Torales Sistemas de Bases de Datos Federadas para la Gestión de la Información. In: JIT-CITA, 2013 Asunción <http://www.cyt.uc.edu.py/jit-cita/2013/index.php/contenido-programa>. 2013.

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Bases de datos Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Bases de datos

Medio: Internet.

Sitio Web: <https://tinyurl.com/SBDF-GI>

Palabras Clave: fuentes de datos heterogeneas; integracion de informacion; gestion de informacion; implementacion de software; bases de datos federadas;

Observaciones: Presentación de avances de trabajos de Grado y Postgrado

Resúmenes simples en anales de eventos

1 Marvin M. Agüero-Torales Gastro-miner. In: I Workshop de Ciencias de Datos - Universidad Comunera, 2018 Asunción <https://www.ucom.edu.py/wp-content/uploads/2019/08/Jornadas-Cientificas-en-Ciencia-de-Datos-UCOM.pdf>. 2018.

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Minería de opiniones

Medio: Otros.

Sitio Web: <http://dx.doi.org/10.13140/RG.2.2.33136.71688>

Palabras Clave: minería de opinión; análisis de sentimientos; redes sociales;

Observaciones: Póster científico ganador del concurso I-Data: <https://www.ucom.edu.py/cienciadedatos/concurso-i-data/resultados/>

Resúmenes expandidos en anales de eventos

1 RUBEN DIAZ; DIEGO DIAZ; EMILIO DIAZ; MARGARITA RUIZ-OLAZAR; Marvin M. Agüero-Torales; Building a Large Language Model for Guarani-Jopara? Methodology, Challenges, and Preliminary Results. In: Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP), 2025 Albuquerque, Ciudad en Nuevo México, USA NON-ARCHIVAL. 2025.

Medio: Otros.

Sitio Web: https://turing.iimas.unam.mx/americasnlp/2025_workshop.html

Observaciones: NON-ARCHIVAL EXTENDEN RESUME PRESENTED AS POSTER.

Building a Large Language Model for Guarani-Jopara. Methodology, Challenges, and Preliminary Results
May 2025

DOI: 10.13140/RG.2.2.32323.52006

Conference: Fifth Workshop on NLP for Indigenous Languages of the Americas (AmericasNLP) co-located at NAACL 2025

<http://dx.doi.org/10.13140/RG.2.2.32323.52006>

2 RUBEN DIAZ; DIEGO DIAZ; Marvin M. Agüero-Torales; MARGARITA RUIZ-OLAZAR; Modelo Grande de Lenguaje para Guaraní - Jopará: Evaluación de Factibilidad y Metodología. In: II Congreso Paraguayo de Informática 2024, 2024 Encarnación, Itapúa 2024.

Medio: Otros.

Sitio Web: <https://congreso-paraguay-en-informatica.github.io/ConParInf2024/index.html#schedule>

Observaciones: Modelo Grande de Lenguaje para Guaraní - Jopará: Evaluación de Factibilidad y Metodología

Rubén Díaz, Diego Díaz, Marvin Agüero-Torales, Margarita Ruiz Olazar

<https://docs.google.com/spreadsheets/d/1VS9nkVV1wrWiAUnZFRreEpX1WG0VPjhnKX4na4BgdIIA/edit?gid=1378177242#gid=1378177242>
(PARTE DE LA TESIS DE MAESTRIA EN CIENCIAS DE DATOS EN LA UCOM)

3 Marvin M. Agüero-Torales Avances y Perspectivas Futuras del Procesamiento de Lenguaje Natural para el Idioma Guaraní. In: II Congreso Paraguayo de Informática 2024, 2024 Encarnación, Itapúa 2024.

Medio: Otros.

Observaciones: Trabajo Invitado: Avances y Perspectivas Futuras del Procesamiento de Lenguaje Natural para el Idioma Guaraní

<https://docs.google.com/spreadsheets/d/1VS9nkVV1wrWiAUnZFRreEpX1WG0VPjhnKX4na4BgdIIA/edit?gid=271629872#gid=271629872>

4 Marvin M. Agüero-Torales; Laura Fólica; La traducción literaria en revistas digitalizadas iberoamericanas (1898-1959): ventajas e inconvenientes del trabajo con herramientas de las HD. In: V congreso de la HDH 2021, 2021 Santiago de Compostela (Online) <https://hdh2021.org/assets/hdh/Libro de resúmenes HDH 2021 - completo.pdf>. 2021.

Medio: Internet.

Sitio Web: <https://hdh2021.org/>

Observaciones: Redes Sociales del Pasado. Cartografía de la Modernidad Literaria Hispánica, 1898-1959

Modera: Diana Roig Sanz · Asiste: César González-Pérez

Panelistas: Ainamar Clariana; Aurea Mota; Elisabet Carbó; Laura Fólica; Marc Moreno Galimany; Marvin Matías Agüero Torales; Pablo Suárez-Mansilla; Ventsi Ikoff

5 Marvin M. Agüero-Torales; Laura V. Vigoya; Elisa Fernández-Álvarez; Susana Ladra; Aplicación de algoritmos de machine learning para peritaje vehicular. In: Machine Learning Workshop Galicia 2019: WGML2019, 2019 A Coruña Actas del IV Machine Learning Workshop Galicia. 2019.

Medio: Internet.

Sitio Web: <https://hdh2021.org/assets/hdh/Libro de resúmenes HDH 2021 - completo.pdf>

Artículos publicados en revistas científicas

Artículos completos publicados en revistas arbitradas

- Marvin M. Agüero-Torales; Carlos Rodríguez Abellán; Marta Carcajona Mata; Juan I. Díaz Hernández; Mario Solís López; Antonio Miranda-Escalada; Sergio López-Alvárez; Jorge Mira Prats; Carlos A. Castaño Moraga; DAVID VILARES; Luis Chiruzzo; Overview of GenoVarDis at IberLEF 2024: NER of Genomic Variants and Related Diseases in Spanish, *Procesamiento del Lenguaje Natural*, 2024.**

Medio: Internet. ISSN/ISBN: 1989-7553

Sitio Web: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6625>

Observaciones: Agüero-Torales, M., Rodríguez Abellán, C., Carcajona Mata, M., Díaz Hernández, J., Solís López, M., Miranda-Escalada, A., López-Alvárez, S., Mira Prats, J., Castaño Moraga, C., Vilares, D., & Chiruzzo, L. (2024). Overview of GenoVarDis at IberLEF 2024: NER of Genomic Variants and Related Diseases in Spanish. *Procesamiento Del Lenguaje Natural*, 73, 381-392. Recuperado de <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6625>

- Luis Chiruzzo; Marvin M. Agüero-Torales; Gustavo Giménez-Lugo; Aldo Alvarez; Yliana Rodriguez; Santiago Góngora; Tamar Solorio; Overview de GUA-SPA en IberLEF 2023: Análisis de Cambio de Código en Guaraní-Español, *Procesamiento del Lenguaje Natural*, 2023.**

Medio: Internet. ISSN/ISBN: 1989-7553

Sitio Web: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6563>

Palabras Clave: ner; code-switching; jopará; guaraní; language identification; shared-tasks;

Observaciones: Chiruzzo, L., Agüero-Torales, M., Giménez-Lugo, G., Alvarez, A., Rodríguez, Y., Góngora, S., & Solorio, T. (2023). Overview of GUA-SPA at IberLEF 2023: Guaraní-Spanish Code Switching Analysis. *Procesamiento Del Lenguaje Natural*, 71, 321-328. Recuperado de <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6563/3963>

- Marvin M. Agüero-Torales; A.G. López-Herrera; DAVID VILARES; (RELEVANTE) Multidimensional Affective Analysis for Low-Resource Languages: A Use Case with Guaraní-Spanish Code-Switching Language, *Cognitive Computation*, 2023.**

Medio: Otros. ISSN/ISBN: 1866-9956

Sitio Web: <https://link.springer.com/article/10.1007/s12559-023-10165-0#citeas>

- Marvin M. Agüero-Torales; JOSE I. ABREU SALAS; A.G. López-Herrera; (RELEVANTE) Deep learning and multilingual sentiment analysis on social media data: An overview, *Applied Soft Computing Journal*, v. 107, 2021.**

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática /

Medio: Internet. ISSN/ISBN: 1568-4946

Sitio Web: <https://www.sciencedirect.com/science/article/abs/pii/S1568494621002969>

Observaciones: Part of special issue:

SI: Soft Computing for Recommender Systems and Sentiment Analysis

Edited by Lorenzo Malandri, Carlos Porcel, Frank Xing, Jesus Serrano-Guerrero, Erik Cambria

- Marvin M. Agüero-Torales; DAVID VILARES; A.G. López-Herrera; (RELEVANTE) Discovering topics in Twitter about the COVID-19 outbreak in Spain, *Procesamiento del Lenguaje Natural*, v. 66, p. 177-190, 2021.**

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / PROCESAMIENTO DE LENGUAJE NATURAL

Medio: Internet. ISSN/ISBN: 1989-7553

Sitio Web: <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6333>

Observaciones: Parte del PLAN-TL

- Marvin M. Agüero-Torales; M. J. Cobo; ENRIQUE HERRERA-VIEDMA; A.G. López-Herrera; (RELEVANTE) A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor, *Procedia Computer Science*, v. 162, p. 392-399, 2019.**

Medio: Internet. ISSN/ISBN: 1861-1870

Sitio Web: <https://www.sciencedirect.com/science/article/pii/S1877050919320125>

Palabras Clave: opinion mining; sentiment analysis; tripadvisor; software tool; cloud;

Artículos completos publicados en revistas NO arbitradas

- Marvin M. Agüero-Torales; M. Canto; Jose Luis Vazquez Noguera; (RELEVANTE) Sistemas de Bases de Datos Federadas para la Gestión de la Información, *Researchgate.net*, 2016.**

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Integración de datos

Medio: Internet. ISSN/ISBN: 0000-0000

Sitio Web: <http://dx.doi.org/10.13140/RG.2.2.29045.88802>

Artículos resumidos publicados en revistas

- Marvin M. Agüero-Torales (RELEVANTE) Machine Learning approaches for Topic and Sentiment Analysis in multilingual opinions and low-resource languages: From English to Guaraní, *Procesamiento del Lenguaje Natural*, 2023.**

Medio: Otros. ISSN/ISBN: 1989-7553

Libros y capítulos de libros publicados

Libros publicados

- 1 **Marvin M. Agüero-Torales** **Machine Learning approaches for Topic and Sentiment Analysis in multilingual opinions and low-resource languages: From English to Guarani (en prensa), Granada, Universidad de Granada. Tesis Doctorales, 2022, v. 1, p. 249, ISSN/ISBN: 9788411172332**

Medio: Internet.

ISSN/ISBN: 9788-4111

Sitio Web: <http://hdl.handle.net/10481/72863>

Observaciones: Esta tesis se ha centrado en el estudio de técnicas de aprendizaje automático para el análisis de sentimientos y el modelado de temas en textos procedentes de medios sociales. Se ha puesto un énfasis especial en los enfoques y métodos para el manejo de idiomas conocidos como low-resource.

Documentos de trabajo

- 1 **Marvin M. Agüero-Torales; Eulàlia Farré-Maduell; Martin Krallinger; Antonio Miranda-Escalada; Salvador Lima-López; Vicent Briva-Iglesias; ProfNER guidelines - English version, 2021.**

Medio: Internet.

Sitio Web: <http://dx.doi.org/10.5281/zenodo.4479741>

Observaciones: SMM4H 2021 accepted at NAACL (scheduled in Mexico City in June) <https://2021.naacl.org/>. These guidelines describe the process followed by the clinical and linguist experts who manually annotated the ProfNER corpus. English version of ProfNER annotation guidelines. The Spanish version here, <https://doi.org/10.5281/zenodo.4306016>. For further information, please visit <https://temu.bsc.es/smm4h-spanish/>

- 2 **Marvin M. Agüero-Torales** **Guarani and Jopara sentiment analysis - Mini-annotation guides, 2021.**

Medio: Internet.

Sitio Web: <http://dx.doi.org/10.13140/RG.2.2.21174.63048>

Observaciones: An annotation mini-guidelines, which describes the process followed by the bilingual annotators (Guarani-Spanish) who manually annotated the Guarani-dominant Jopara and Guarani corpus (<https://github.com/mmaguero/josa-corpus>)

- 3 **Marvin M. Agüero-Torales** **Quantitative analysis for Topic Modeling - Mini-evaluation framework and mini-guides, 2021.**

Medio: Internet.

Sitio Web: <http://dx.doi.org/10.13140/RG.2.2.18639.20647>

Observaciones: Mini-evaluation framework and mini-guides. See more details in (section 4.4, p. 186, <http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6333>).

- 4 **Marvin M. Agüero-Torales** **Emotion detection, offensive language identification, and humor detection in Guarani and Jopara - Mini-annotation guides, 2021.**

Medio: Internet.

Sitio Web: <http://dx.doi.org/10.13140/RG.2.2.12786.02244>

Observaciones: An annotation mini-guidelines, which describes the process followed by the bilingual annotators (Guarani-Spanish) who manually annotated the Guarani-dominant Jopara and Guarani corpus (<https://github.com/search?q=user:mmaguero+corpus>)

- 5 **Marvin M. Agüero-Torales** **Disertación de Máster-Gastro-miner: Una Herramienta Basada en la Nube para el Análisis de Sentimientos en Opiniones sobre Restaurantes en TripAdvisor: Caso de Estudio sobre Restaurantes de la Provincia de Granada, 2017.**

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Ciencias de datos Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Redes Sociales Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Cloud Computing Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Ingeniería web

Medio: Otros.

Sitio Web: <http://etsiit.ugr.es/pages/biblioteca>

Palabras Clave: tripadvisor; minería de opiniones; análisis de sentimientos; cloud computing; herramienta software; medios sociales;

Observaciones: Disertación de Máster en Ingeniería Informática, Universidad de Granada - 2017.

Tutor: Antonio G. López-Herrera

Co-tutor: Manuel Jesús Cobo

- 6 **Marvin M. Agüero-Torales** **Disertación de Grado-Sistemas de Bases de Datos Federadas para la Gestión de la Información, 2015.**

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Bases de datos Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Bases de datos

Medio: Papel.

Sitio Web: <http://www.uaa.edu.py>

Palabras Clave: fuentes de datos heterogeneas; gestion de informacion; integracion de informacion; bases de datos federadas; implementacion de software;

Observaciones: Disertación de Grado en Ingeniería en Sistemas Informáticos, Universidad Autónoma de Asunción (UAA), 2015.

Tutor: Mario Rubén Canto.

Evaluaciones

Evaluación de Eventos

IberLEF 2025 Shared-tasks (España) (2025)

Observaciones: Review shared-tasks proposals

IberLEF 2024 Shared-tasks (España) (2024)

Observaciones: Review shared-tasks proposals

Evaluación de Publicaciones

IberLEF 2024 Ceur-WS (2024 - 2024)

Cantidad: Menos de 5. Observaciones: Edición y revisión del track: "BIOMEDICAL NLP - NER in Genomic Variants and related Diseases in Spanish (GenoVarDis)"

URL: <https://ceur-ws.org/Vol-3756/>

IberLEF 2023 Ceur-WS (2023 - 2023)

Cantidad: Menos de 5. Observaciones: CODE SWITCH ANALYSIS

Guarani-Spanish Code-Switching Analysis (GUA-SPA)

<https://ceur-ws.org/Vol-3496/>

Tutorías/Orientaciones/Supervisiones

En Marcha

Tesis de maestría

1 Diego Díaz, Ruben Díaz, - Cotutor o Asesor - Factibilidad de gran modelo de lenguaje para el Guaraní/Jopará, 2025

Disertación (Maestría en Ciencias de Datos), UCOM - Universidad Comunera, Paraguay

País: Paraguay / Idioma: Español

Palabras Clave: ia; nlp; llm; jopará; guaraní; fine-tuning; low-resources languages;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y Procesamiento de Lenguaje Natural

Observaciones: con varios trabajos publicados en congresos nacionales, internacionales y regionales. Tutoreado junto Margarita Ruíz-Olazar

Concluidas

Otras tutorías/orientaciones

1 Cecilia Gonzáles, Ramón Araujo, - Cotutor o Asesor - Pipeline de limpieza para corpus en guaraní, 2025

Otras tutorías/orientaciones (Ingeniería Informática), UCA - Universidad Católica "Nuestra Señora de la Asunción", Paraguay

País: Paraguay / Idioma: Español

Palabras Clave: nlp; corpus; dataset; pipeline; corpora; language identification; deduplication; ingestion;

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / Procesamiento de Lenguaje Natural

Observaciones: Autumn code UC 2025 de la asignatura de Ing. de Software (responsable: Luca Cernuzzi)

Otras Referencias

Premiaciones

- 1 Best paper award, NAACL (North American Chapter for Computational Linguistics), 2024: Theme track: Native American Languages -Indigenous Languages (internacional), Association for Computational Linguistics (2024)**
 Best paper award, NAACL (North American Chapter for Computational Linguistics), 2024: Theme track: Native American Languages -Indigenous Languages-, "Grammar-based Data Augmentation for Low-Resource Languages: The Case of Guarani-Spanish Neural Machine Translation" (<https://aclanthology.org/2024.naacl-long.354>).
- 2 CUM LAUDE PHD THESIS (nacional), Universidad de Granada (2022)**
 Thesis Cum Laude Mention, in Doctoral Program in ICTs, University of Granada, Spain - 2022: Thesis distinction.
- 3 Primer puesto concurso I-Data (nacional), Universidad Comunera (2018)**
 Ganador del póster científico
- 4 Subvención de Matrícula - Doctorado en TIC (nacional), Universidad de Granada (2017)**
 Beneficiario de una subvención de matrícula para el Programa de Doctorado en TICs
- 5 BeCAL Segunda Convocatoria (nacional), Becas de postgrado en el exterior Don Carlos Antonio López (2016)**
 Becado a España para realizar un Máster en Ing. Informática en la Universidad de Granada
- 6 Mención de Honor Licenciatura en Ciencias Informáticas (nacional), Universidad Autónoma de Asunción (2011)**
 Licenciatura en Ciencias Informáticas, Graduado con Honores, Especialización en Sistemas Informáticos
- 7 Scholarship Program of the Faculty of Sciences and Technologies of the Autonomous University of Asunción (nacional), Universidad Autónoma de Asunción (2007)**
 Beneficiary of the Scholarship Program of the Faculty of Sciences and Technologies of the Autonomous University of Asunción, Paraguay; 2006, full scholarship; 2007/09, half scholarship: Better academic record.
- 8 Program of Support to Better Alumni (P.A.M.E.) (nacional), Municipalidad de Asunción (2006)**
 Beneficiary of the Program of Support to Better Alumni (P.A.M.E.) of the Municipality of Asunción, Paraguay - 2006/10: Awarded for being one of the best graduates of a public institution in Asunción. Financial aid for university studies.
- 9 Medalla de Oro Bachillerato en Ciencias Básicas (nacional), Colegio Nacional EMD Pdte Franco (2005)**
 Mejor egresado de la Promoción 2005
- 10 Medalla de Oro Escolar Básica (nacional), Escuela Nacional Gral. José Elizardo Aquino (1999)**
 Mejor egresado de la Promoción 1999

Presentaciones en eventos

- 1 Taller - GenoVarDis at IberLEF 2024: NER of Genomic Variants and Related Diseases in Spanish, 2024, España**
 Nombre: Iberian Languages Evaluation Forum. Tipo de Participación: Expositor oral - Información Adicional: We present the first shared task for Named Entity Recognition (NER) in Spanish written scientific literature about genomic variants, genes, and its associated diseases and symptoms, GenoVarDis at IberLEF 2024 campaign. The challenge consisted on identifying entities related to genomic variants. We annotated a corpus of 633 abstracts extracted from PubMed articles, with the information for the tasks. Seven teams took part in the evaluation phase, obtaining in general good results for the task.
 Nombre de la institución promotora: Sociedad Española para el Procesamiento del Lenguaje Natural
 Áreas de Conocimiento:
 Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / IA y PROCESAMIENTO DE LENGUAJE NATURAL
- 2 Taller - On the logistical difficulties and findings of Jopara Sentiment Analysis, 2021, México**
 Nombre: CALCS 2021 (co-located with NAACL 2021) - Fifth Workshop on Computational Approaches to Linguistic Code Switching (June 2021). Tipo de Participación: Expositor oral - Información Adicional: Corresponding to the paper: <https://aclanthology.org/2021.calcs-1.12/>
 Nombre de la institución promotora: Association for Computational Linguistics
 Áreas de Conocimiento:
 Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / NATURAL LANGUAGE PROCESSINGHumanidades / Lengua y Literatura / Lingüística / COMPUTATIONAL LINGUISTICS
- 3 Congreso - La traducción literaria en revistas digitalizadas iberoamericanas (1898-1959): ventajas e inconvenientes del trabajo con herramientas de las HD, 2021, España**
 Nombre: V Congreso de la Sociedad Internacional de Humanidades Digitales Hispánicas. Tipo de Participación: Panelista - Información Adicional: Tareas desarrolladas en el marco del proyecto ERC: Redes Sociales del Pasado. Cartografía de la Modernidad Literaria Hispánica, 1898-1959.
 Nombre de la institución promotora: HUMANIDADES DIGITALES HISPANICAS
 Áreas de Conocimiento:
 Humanidades / Otras Humanidades / DIGITAL HUMANITIES
- 4 Otra - #HackHealthBCN, 2020, España**
 Nombre: <https://hackathonspain.com/hackhealthbcn-2020>. Tipo de Participación: Otros - Información Adicional: #HackHealthBCN tiene como objetivo acercar los desafíos reales a los que se enfrentan los profesionales del campo de la salud en su día a día a la comunidad tecnológica, desarrollando e impulsando el sector de la bioinformática.

 Durante los tres días los participantes desarrollarán sus soluciones sobre la plataforma IBM Power Systems para inteligencia artificial y Mainframe para blockchain

RETO 2: Casos clínicos

El objetivo de este reto es fomentar el desarrollo de sistemas de minería de textos para poder encontrar similitudes entre una colección de

textos de casos clínicos pasados escritos en español.

Nombre de la institución promotora: BIOINFORMATICS BARCELONA

5 Taller - Aplicación de algoritmos de machine learning para peritaje vehicular, 2019, España

Nombre: Machine Learning Workshop Galicia 2019. Tipo de Participación: Expositor oral - Información Adicional: Resumen presentado como poster: «Aplicación de algoritmos de machine learning para peritaje vehicular»

Marvin M. Agüero-Torales, Laura V. Vigoya, Elisa Fernández Álvarez and Susana Ladra

Nombre de la institución promotora: Universidad de la Coruña

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática /

6 Congreso - Concurso I-Data, 2018, Paraguay

Nombre: Workshop en Ciencias de Datos. Tipo de Participación: Poster - Información Adicional: Póster ganador

Nombre de la institución promotora: Universidad Comunera

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Minería de opiniones

7 Congreso - Sistemas de Bases de Datos Federadas para la Gestión de la Información, 2013, Paraguay

Nombre: JIT-CITA. Tipo de Participación: Expositor oral - Información Adicional: Concurso de Trabajos estudiantiles, de grado y postgrado

Nombre de la institución promotora: Universidad Autonoma de Asunción

Áreas de Conocimiento:

Ingeniería y Tecnología / Ingeniería Eléctrica, Ingeniería Electrónica e Ingeniería de la Información / Ingeniería de Sistemas y Comunicaciones / Bases de datos Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Computación / Bases de datos

Jurado/Integrante

Otros tipos

- 1 Marvin M. Agüero-Torales; Jorge Bendahan; Hugo Correa; Participación en comités de Antoliano Cáceres Estigarribia. Tesis/Monografía de grado Análisis de la Sensibilidad de los Algoritmos Naive Bayes y Regresión Logística con Respecto a Atributos del Problema y Tamaño de la Muestra, 2019, Paraguay/Español**

Otra participación (Ingeniería en Sistemas Informáticos), UAA - Universidad Autonoma de Asunción

Áreas de Conocimiento:

Ciencias Naturales / Ciencias de la Computación e Información / Ciencias de la Información y Bioinformática / machine learning

Obs: En Machine Learning la tareade automatizar el aprendizaje de máquinas es un campo en constante desarrollo y su valor va en aumento constante por los aportes en conocimiento que dan a diversas áreas de la actividad humana, como por ejemplo la salud.En este trabajo presentamos un estudio sobre dos algoritmos de inteligencia artificial por ser los más utilizados actualmente para modelar los datos, Naive Bayes y Regresión Logística. Se propone una metodología que mide el error de predicción de ambos algoritmos, estimando la sensibilidad del error con respecto a dimensión del problema hasta 6 atributos y tamaño de la muestra de entrenamientohasta 500 instancias. Para la evaluación de cada algoritmo se utiliza una distribución aleatoria del cual también es generada aleatoriamente cada muestra de entrenamiento. Con los experimentos realizados hemos observadosobreCuán sensible es el error al tamaño del problema y de la muestrapara cada algoritmo.Teniendo en cuenta las características de los datos y de las variables estudiadas observamos que la mayor sensibilidad del error se observa hasta 300 instancias deentrenamiento y a partir de ahí la sensibilidad tiende a minimizarse.El algoritmo de Regresión Logística presenta una superioridad con respecto a Naive Bayes en mínimas proporciones.

Información adicional:

Indicadores

Producción Técnica

26

Cursos de corta duración dictados

3

Otro

1

Especialización

1

Perfeccionamiento

1

Procesos o técnicas	5
Otros procesos o técnicas	5
Productos tecnológicos	18
Otro	1
Software	17
Producción Bibliográfica	29
Trabajos en eventos	14
Completo	8
Resumen expandido	5
Resumen	1
Artículos publicados en revistas científicas	8
Completo en revistas arbitradas	6
Completo en revistas NO arbitradas	1
Resumen	1
Libros y capítulos de libros publicados	1
Libro publicado	1
Documentos de trabajo	6
Completo	6
Tutorías	2
En Marcha	1
Tesis de maestría	1
Concluidas	1
Otras tutorías/orientaciones	1
Evaluaciones	4
Eventos	2
Publicaciones/Periódicos	2
Otras Referencias	18
Otros datos Relevantes	10
Presentaciones en eventos	7
Jurado/Integrante	1