

Proyecto de grado

...

2023

Guaraní



Lengua de origen indígena que se habla masivamente en 🇵🇷 y en regiones de 🇧🇷 🇦🇷 🇨🇴

Aunque se habla muchísimo no se usa tanto **formalmente*** en las redes y la web (difícil recolectar datos)

Guaraní :: Escasos recursos

Datos

“Crawlers” web



Identificadores
de lengua

Extracción de
información

Guaraní :: Escasos recursos


Sin embargo:

DETECTAR IDIOMA INGLÉS ESPAÑOL GUARANÍ

Hola, mi nombre es Juan y tengo un perro.

Maitei, che réra Juan ha areko peteĩ jagua.

41 / 5.000




DETECTAR IDIOMA INGLÉS ESPAÑOL GUARANÍ

Ko'ẽrõ oñepresenta Guaraní Ñe'ẽtekuaa.

Mañana se presentan los Guardianes de la Galaxia.

38 / 5.000



Guaraní :: En qué hemos trabajado

Development of a Guaraní - Spanish Parallel Corpus

Luis Chiruzzo¹, Pedro Amarilla², Adolfo Ríos², Gustavo Giménez Lugo³

¹Universidad de la República, Montevideo, Uruguay

²Universidad Nacional de Asunción, San Lorenzo, Paraguay

³Universidade Tecnológica Federal do Paraná, Curitiba, PR - Brasil

luischir@fing.edu.uy, pj.amarilla@gmail.com, adolforinu@gmail.com, gustavo@dainf.ct.utfpr.edu.br

Experiments on a Guaraní Corpus of News and Social Media

Santiago Góngora, Nicolás Giossa, Luis Chiruzzo

Facultad de Ingeniería

Universidad de la República

Montevideo, Uruguay

{sgongora,nicolas.giossa,luischir}@fing.edu.uy

Jojajovai: A Parallel Guaraní-Spanish Corpus for MT Benchmarking

**Luis Chiruzzo¹, Santiago Góngora¹, Aldo Alvarez²,
Gustavo Giménez-Lugo³, Marvin Agüero-Torales^{4,5}, Yliana Rodríguez¹**

¹Universidad de la República, Montevideo, Uruguay

²Universidad Nacional de Itapúa, Encarnación, Paraguay

³Universidade Tecnológica Federal do Paraná, Curitiba, PR - Brasil

⁴Universidad de Granada, Granada, Spain

⁵Internet Interdisciplinary Institute (IN3), Universitat Oberta de Catalunya, Barcelona, Spain

luischir@fing.edu.uy, sgongora@fing.edu.uy, aldo.alvarez@fiuni.edu.py,

gustavogl@utfpr.edu.br, maguero@correo.ugr.es, yrodriguez@fhuce.edu.uy

Can We Use Word Embeddings for Enhancing Guaraní-Spanish Machine Translation?

Santiago Góngora, Nicolás Giossa, Luis Chiruzzo

Facultad de Ingeniería

Universidad de la República

Montevideo, Uruguay

{sgongora,nicolas.giossa,luischir}@fing.edu.uy

Initial Experiments for Building a Guaraní WordNet

Luis Chiruzzo¹ Marvin M. Agüero-Torales^{2,3} Aldo Alvarez⁴ Yliana Rodríguez¹

¹Universidad de la República, Montevideo, Uruguay

²University of Granada, Granada, Spain

³Global CoE of Data Intelligence, Fujitsu, Madrid, Spain

⁴Universidad Nacional de Itapúa, Encarnación, Paraguay

luischir@fing.edu.uy, maguero@correo.ugr.es, aldo.alvarez@fiuni.edu.py, yrodriguez@fhuce.edu.uy

Using Guaraní Verbal Morphology on Guaraní-Spanish Machine Translation Experiments

Uso de la Morfología Verbal del Guaraní en Experimentos de Traducción Automática Guaraní-Español

Yanina Borges, Florencia Mercant, Luis Chiruzzo

Universidad de la República

Montevideo, Uruguay

{yanina.borges,mijailovich, florencia.mercant, luischir}@fing.edu.uy

Guaraní :: Posibles líneas de trabajo

Trabajar a fondo en un sistema de recolección de datos de la web

Diseñar una gramática que genere ejemplos sencillos de oraciones (y usarlas para entrenar)

Evaluar *word embeddings* nuevamente (más datos y otros algoritmos)

Otras ideas (más lingüísticas)
POS tagging
NER
Parsing

