



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA



**DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN
CARRERA DE SOFTWARE**

**ARTÍCULO ACADÉMICO PREVIO A LA OBTENCIÓN DEL TÍTULO
DE INGENIERO EN SOFTWARE**

**TEMA: “ETIQUETADOR AUTOMÁTICO WEB DE ENTIDADES BIOMÉDICAS EN IDIOMA
ESPAÑOL ”**

**AUTORES:
MOREJÓN CORRALES, SAMIEL ANDRÉS
IZA IZA, JOSSELYN LIBELIA**

**DIRECTOR:
ING. UYAGUARI UYAGUARI, ALVARO
DANILO**

LATACUNGA AGOSTO, 2022



AGENDA

- Resumen
- Antecedentes investigativos
- Planteamiento del problema
- Objetivo general y Específicos
- Sistemas y Métodos
- Resultados
- Conclusiones



ANTECEDENTES INVESTIGATIVOS

El paciente tiene fiebre y dolor alrededor del ombligo

Fiebre = Gripe Normal

**Fiebre y Dolor en el ombligo =
Apendicitis**



ANTECEDENTES INVESTIGATIVOS

Sistema Inteligente para realizar predicciones medicas

Tomar amoxicilina cada 2 horas
Ingerir paracetamol después de cada comida
Consumir buscapina dos veces al día

Fractura de Brazo

Diccionarios
(UMLS)



ANTECEDENTES INVESTIGATIVOS

Sistema Inteligente para realizar predicciones medicas

Tomar amoxicilina cada 2 horas
Ingerir paracetamol después de cada comida
Consumir buscapina dos veces al día

Ingerir aspirina todas las mañanas

Fractura de Brazo

Diccionarios
(UMLS)



ANTECEDENTES INVESTIGATIVOS

Sistema Inteligente para realizar predicciones medicas

Tomar amoxicilina cada 2 horas
Ingerir paracetamol después de cada comida
Consumir buscapina dos veces al día

Ingerir aspirina todas las mañanas

Fractura de Brazo

Fractura = Traumatismo

Brazo = Parte del cuerpo

Diccionarios
(UMLS)



ANTECEDENTES INVESTIGATIVOS



Español

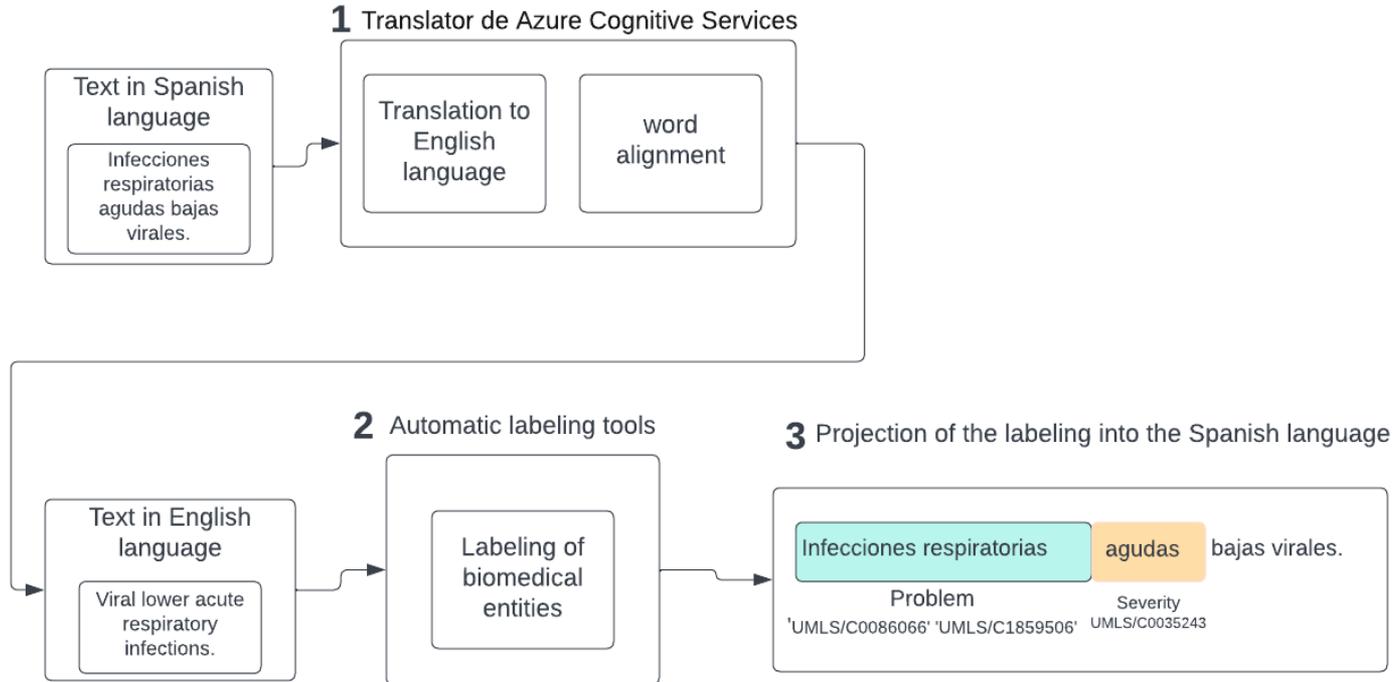


Ingles



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

ANTECEDENTES INVESTIGATIVOS



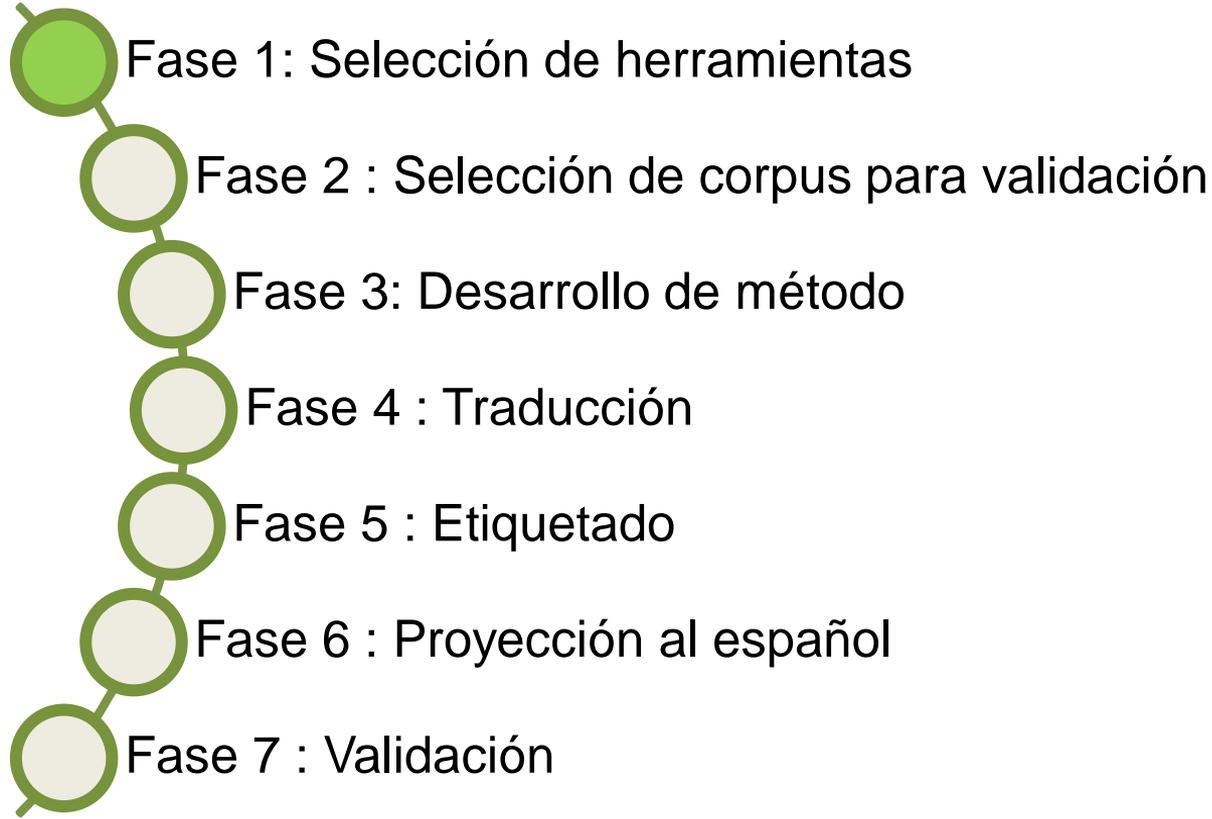
OBJETIVOS GENERAL

Desarrollar un sistema de reconocimiento de entidades biomédicas en textos médicos en lenguaje español, aplicando técnicas de cross-lingual y de procesamiento de lenguaje natural.

OBJETIVOS ESPECÍFICOS

Desarrollar un sistema web para el reconocimiento de entidades biomédicas.





ANTECEDENTES INVESTIGATIVOS

Herramientas de etiquetación automática

Existen herramientas de etiquetación automática que analiza texto médico no estructurado, como registros médicos donde pueden extraer información sobre conceptos médicos, como enfermedades, medicamentos, dispositivos médicos, procedimientos y sus atributos clínicos relevantes.



ANTECEDENTES INVESTIGATIVOS

Herramientas de etiquetación automática

Metamap

Enfoque por búsquedas Heurísticas

Coincidencia de diccionarios para reconocer diversos tipos de conceptos en múltiples fuentes
UMLS, (Centro Nacional de Información Biotecnológica) NCBI, entre otros

API de Healthcare Natural Language

Enfoque de Modelos Supervisados

utiliza algoritmos que aprenden iterativamente de los datos.
Modelos entrenados para extraer conceptos y relaciones de textos médicos



ANTECEDENTES INVESTIGATIVOS

CLEF

El proyecto Clinical E-Science Framework (CLEF) está construyendo un marco para la captura, integración y presentación de información clínica.

EMEA

EMEA es un corpus de documentos biomédicos recuperados de la Agencia Europea de Medicamentos (EMA). Incluye documentos relacionados con medicamentos y sus traducciones a 22 lenguas oficiales de la Unión Europea.

MEDLINE

Medline es una base de datos bibliográficos producida por la Biblioteca Nacional de Medicina de los Estados Unidos. Entre sus ventajas como fuente de información está su amplia cobertura de revistas, pues contiene aproximadamente 15 millones de citas bibliográficas que provienen de más de 4,600 revistas que cubren los temas de la biomedicina, principalmente medicina, enfermería, odontología, oncología, medicina veterinaria, salud pública, ciencias preclínicas y de otras áreas de las ciencias de la vida



ANTECEDENTES INVESTIGATIVOS

Tipos Semánticos	
Identificador	Etiqueta
ANAT	Anatomía Productos químicos y medicamentos
CHEM	Productos químicos y medicamentos
DEVI	Dispositivos
DISO	Trastornos
LIVB	Seres vivos
OBJC	Objetos
PHEN	Fenómenos
PHYS	Fisiología
PROC	Procedimientos



ANTECEDENTES INVESTIGATIVOS

Herramientas evaluadas por grupos semánticos

Herramienta	Grupos Semántico	Porcentaje	Idioma
UMLMapper	CLEF	62%	Español
Freeling-Med	Diagnóstico, enfermedades y medicamentos	90% - 92%	Español
MeSpeN	CLEF	47%	Español
PhamaCoNER	Proteínas, Normalizadas, No - Normalizadas y No claros	86,24%	Español



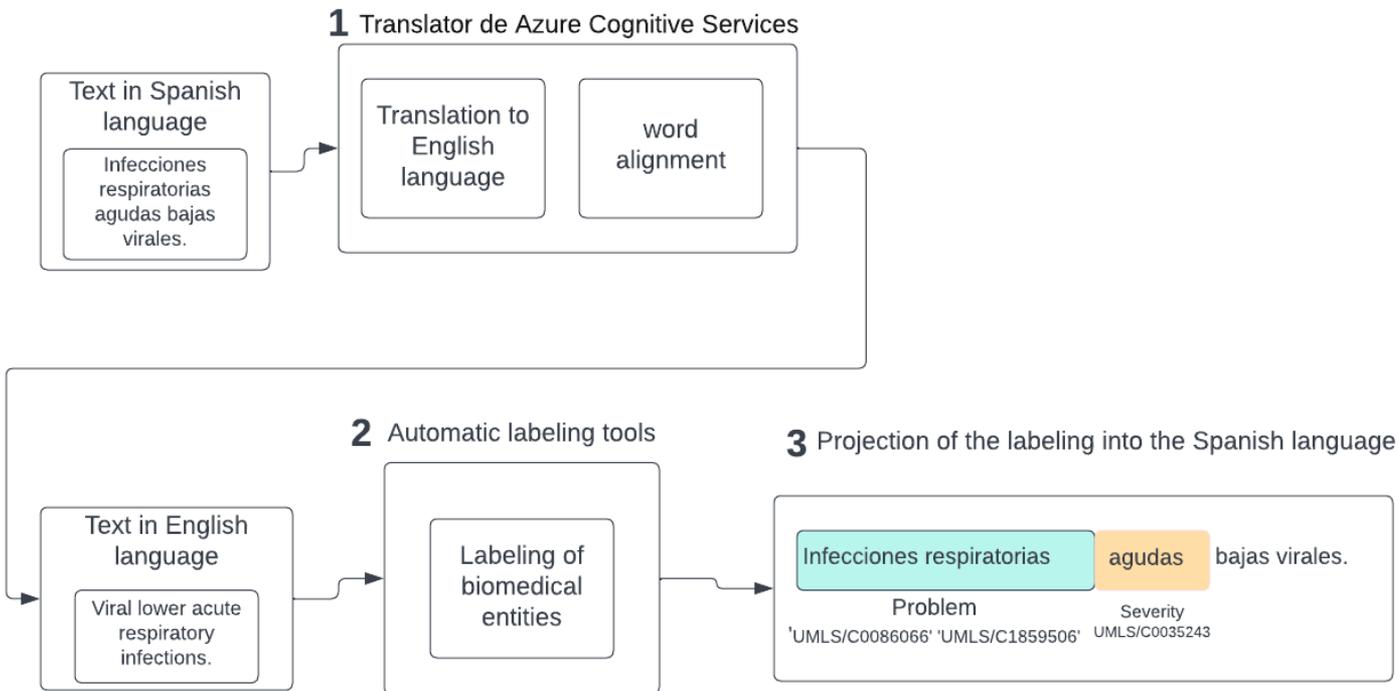
ANTECEDENTES INVESTIGATIVOS

Herramientas evaluadas por grupos semánticos

Herramienta	Grupos Semántico	Porcentaje	Idioma
UMLMapper	CLEF	62%	Español
MeSpeN	CLEF	47%	Español



Transferencia de entidades en el idioma inglés al español



Traducción automática y alineamiento de palabras de entidades en el idioma inglés al español

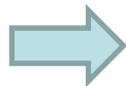
Traductor de Azure Cognitive Services

El servicio Translator es un servicio de traducción automática neuronal basado en la nube que forma parte de la familia de API REST de Azure Cognitive Services

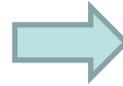
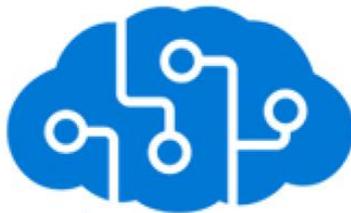


Traducción automática y alineamiento de palabras de entidades en el idioma inglés al español

En pacientes menores de 16 años.



Cognitive Text Translator



In patients under 16 years of age.



Alineamiento	Español	Inglés
0:1-0:1	En	In
3:11-3:10	pacientes	patients
13:19-12:16	menores	under
21:22-27:28	de	of
24:25-18:19	16	16
27:31-21:25	años	years
27:31-30:33	años	Age.



Etiquetado automático de entidades biomédicas

Metamap

Frase	Entidad	Código
The MAH provides educational material to all physicians who may be involved in treating patients with prasugrel.	physicians	C0031831
	prasugrel	C1620287

Api de Healthcare Natural Language

Frase	Entidad	Código
The MAH provides educational material to all physicians who may be involved in treating patients with prasugrel.	prasugrel	C1620287



EJEMPLO

Prevención de la insuficiencia renal más allá del control de la HTA y de la diabetes mellitus.

Traductor

Prevention of renal failure beyond the control of hypertension and diabetes mellitus.

Etiquetador

renal insufficiency	C0035078
HTA	C0020538
Diabetes	C0011849
Diabetes	C0199176
Control	C2587213



EJEMPLO

renal insufficiency	C0035078
HTA	C0020538
diabetes	C0011849
diabetes	C0199176
control	C2587213

Proyección de las entidades al español

renal insuficiencia	C0035078
HTA	C0020538
diabetes	C0011849
diabetes	C0199176
control	C2587213



Validación

Phrase	Entity	Code	Healthcare Natural Language Api	Metamap
Supranuclear gaze palsy following extracorporeal surgery with induced hypothermia. Report of two cases.	Supranuclear gaze palsy	C1720037	1	1
	induced hypothermia	C1301797	0	1
	surgery	C0543467	0	1
	induced hypothermia	C0020674	0	1
	Report	C0700287	0	1

Metamap

Api de Healthcare Natural Language

Metamap puede etiquetar palabras más complejas y frases más extensas , por otra parte la efectividad de Api de Healthcare Natural Language mejora cuando son palabras sencillas y frases más cortas.



Validación

Español

Benign bone tumors of the nose and paranasal sinuses in childhood. Review of cases in the Hospital Infante de México.

Benign bone tumors	C0684516
nose	C0028429
nose	C1278896
paranasal sinuses	C0030471
Hospital	C0019994
México	C0025885

Tumores óseos benignos de nariz y senos paranasales en la infancia. Revisión de casos en el Hospital Infantil de México

nariz	C0028429
nariz	C1278896
senos paranasales	C0030471
Hospital	C0019994
México	C0025885
Tumores óseos benignos	C0684516

Ingles



Sistema Web

Etiquetador de conceptos biomédicos en español

Ingrese su frase

Con la inmunoglobulina humana normal para administración intravenosa pueden producirse, en ocasiones, reacciones adversas como escalofríos, cefalea, fiebre, vómitos, reacciones alérgicas, náuseas, artralgia, hipotensión arterial y lumbalgia moderada.

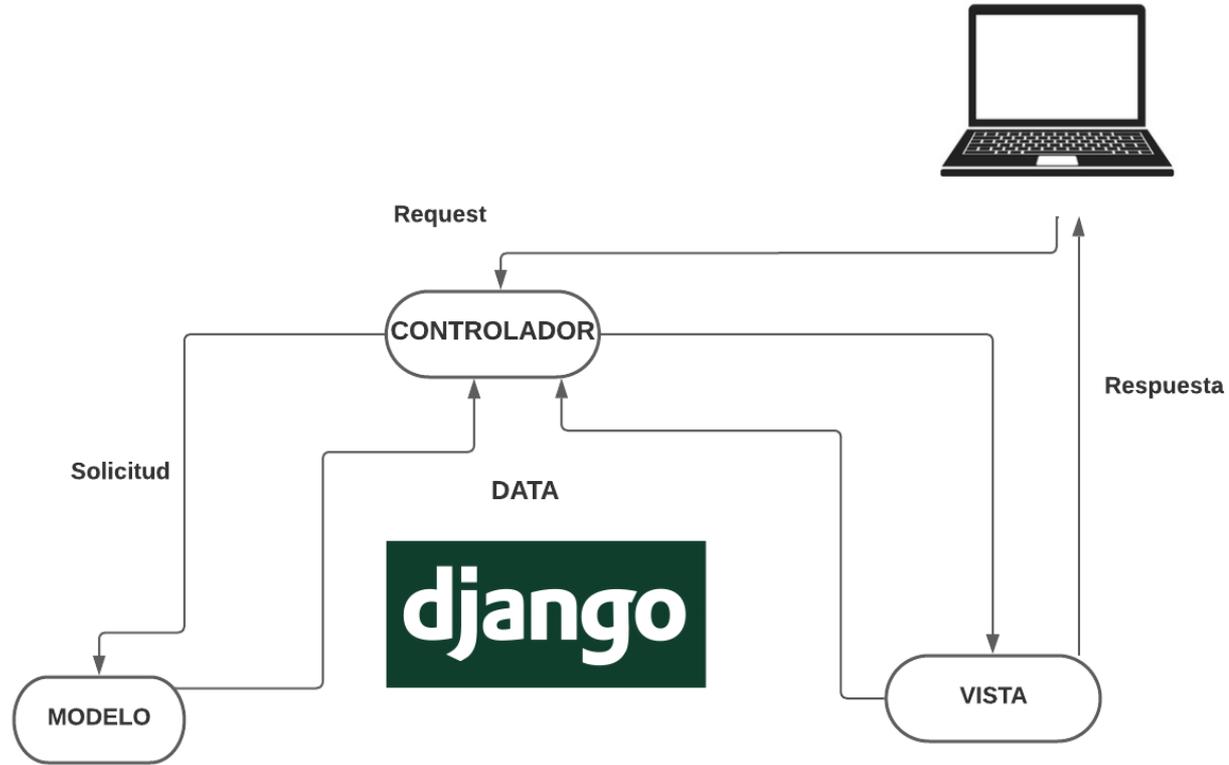
Etiquetar

Tabla

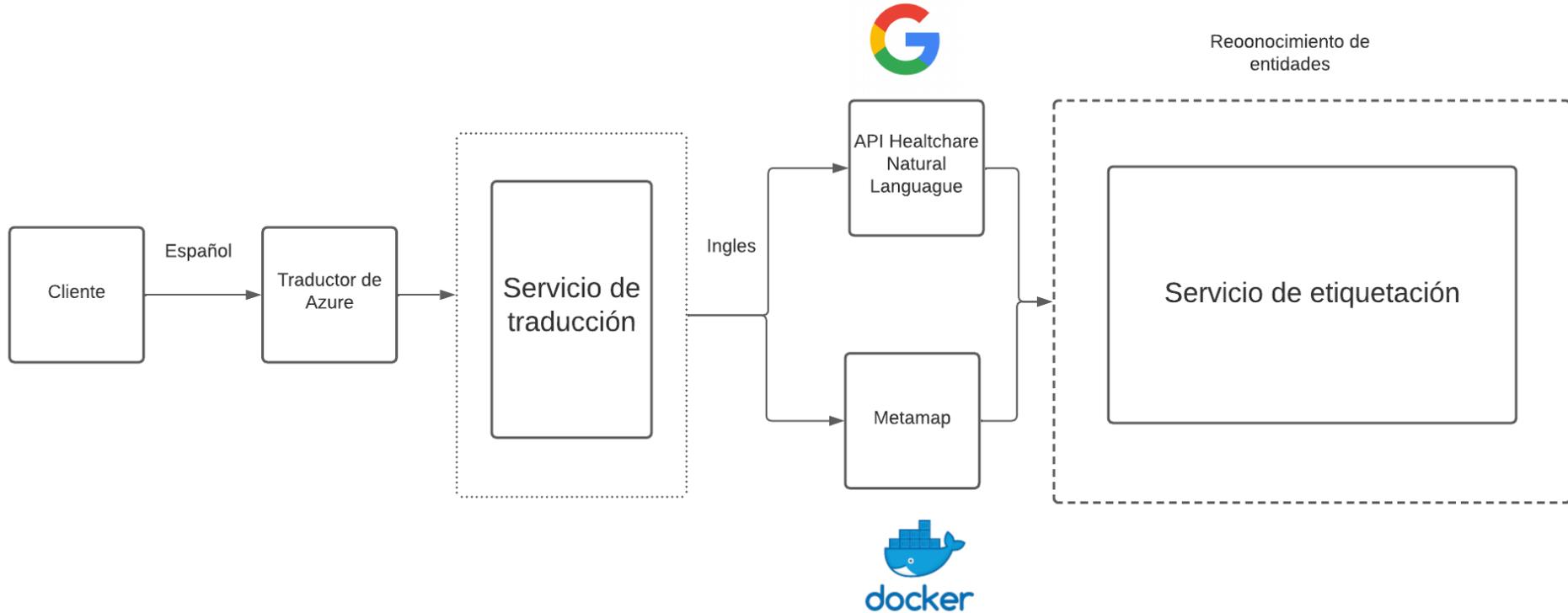
Entidad	Código	Posición Español
reacciones adversas	C0559546	102-120
escalofríos	C0085593	127-138
cefalea	C0015967	140-147
fiebre	C0015967	149-155
vómitos	C0042963	157-164
náuseas	C0027497	188-195
artralgia	C0003862	197-206
lumbalgia	C0024031	231-240
administración intravenosa	C0021440	42-69
reacciones alérgicas	C1527304	166-186



Sistema Web



Sistema Web



Efectividad del etiquetador automático.

Tras haber realizado la validación con las dos secciones del Corpus CLEF, a la aplicación del sistema de reconocimiento y proyección al español obtuvimos los siguientes resultados:

Corpus	Etiquetado con combinación de herramientas	% de efectividad
Medline	166/316	52,53%
EMEA	237/430	55,11%



Efectividad de herramientas de etiquetado.

Corpus	Herramientas de etiquetado de entidades biomédicas	% de efectividad
Medline	Api de Healthcare Natural Language	27,53%
	Metamap	59,49%
EMEA	Api de Healthcare Natural Language	36,04%
	Metamap	56,74%



Efectividad de herramientas de etiquetado.

Api de Healthacare Natural Language

Metamap

Tipos Semánticos

Tipo	Encontrados	Total	% Encontrados	Tipo	Encontrados	Total	% Encontrados
ANAT	17	41	41,46%	ANAT	18	41	43,90%
CHEM	10	28	35,71%	CHEM	16	28	57,14%
DEVI	1	3	33,33%	DEVI	1	3	33,33%
DISO	36	109	33,02%	DISO	73	109	66,97%
LIVB	5	41	12,20%	LIVB	30	41	73,17%
OBJC	0	3	0%	OBJC	3	3	100%
PHEN	1	6	16,66%	PHEN	1	6	16,66%
PHYS	1	15	6,66%	PHYS	9	15	60%
PROC	17	63	26,98%	PROC	31	63	49,20%



Efectividad de herramientas de etiquetado.

Corpus	Herramientas de etiquetado de entidades biomédicas	% de efectividad con traductor	% de efectividad sin traductor
Medline	Api de Healthcare Natural Language	27,53%	30.37%
	Metamap	59.65%	66.45%
EMEA	Api de Healthcare Natural Language	26,27%	38.33%
	Metamap	57,44%	60.50%



Efectividad de herramientas de etiquetado.

Corpus	Herramientas de etiquetado de entidades biomédicas combinadas	% de efectividad con traductor	% de efectividad sin traductor	% de pérdida
Medline	Api de Healthcare Natural Language	65.50%	72.46%	6.69%
	Metamap			
EMEA	Api de Healthcare Natural Language	63.48%	69.53%	6.05%
	Metamap			



Conclusiones

Se puede concluir que usando la herramienta Api de healthcare Natural Language de Google ya teniendo los textos traducidos tanto de corpus Medline y EMEA nos da una efectividad en el etiquetado un 27,53% y 26,27% respectivamente y con la herramienta Metamap se logra un 59,65% y 57,44% correspondientemente.

Se puede demostrar que cada una de estas herramientas tienen un mejor etiquetado en diferentes casos dados en los textos ya que utilizan diferentes enfoques (Búsquedas heurísticas y Modelos supervisados). Es por esto que se realiza la combinación de estas dos herramientas para lograr una mejor efectividad en el proceso de anotación de entidades.

Se presentó la construcción de un etiquetador automático para entidades biomédicas en el idioma español. Utilizando técnicas de cross-lingual para proyectar las entidades identificadas por la combinación de herramientas de anotación biomédicas (Metamap y Api de Healthcare Natural Language) desde el idioma inglés al idioma fuente (español). El mismo que fue validado en los corpus del Medline y EMEA el cual alcanzó un porcentaje 52,53% y 55,11% de efectividad respectivamente, el cual se encuentra en un promedio competitivo con los porcentajes encontrados en los antecedentes investigativos.



Gracias



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA