

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

**TRABAJO DE UNIDAD DE INTEGRACIÓN CURRICULAR PREVIO A LA OBTENCIÓN DEL TÍTULO DE
INGENIERO DE SOFTWARE**

**TEMA: SISTEMA INFORMÁTICO PARA LA NORMALIZACIÓN DE ENTIDADES BIOMÉDICAS
BASADO EN BÚSQUEDAS SEMÁNTICAS MULTILENGUAJE, SOBRE LA BASE DE DATOS MÉDICA
UMLS**

AUTORES:

**BAHAMONDE TONATO, JUAN DIEGO
MONGE AULES, WILLIAM ARIEL**

DIRECTOR:

ING. UYAGUARI UYAGUARI, ALVARO DANILO, Msc

LATACUNGA AGOSTO, 2023



CONTENIDO



Propósito y contextualización del tema

El presente proyecto de investigación busca desarrollar un sistema informático a través de una búsqueda semántica multilinguaje mediante nuevos métodos para la identificación de entidades biomédicas en español utilizando técnicas de procesamiento de lenguaje natural y el paradigma de Cross-lingual.



Justificación del interés de la investigación

El reconocimiento y la normalización de entidades biomédicas en notas clínicas, es un recurso fundamental para encontrar rasgos en el texto y con estos realizar predicciones de diagnósticos y tratamientos médicos, por lo que encontrar nuevos métodos que no requieran corpus etiquetados para realizar la identificación y normalización de entidades biomédicas es de mucha importancia en el área de la bioinformática y la analítica de datos.



CONTENIDO



OBJETIVOS ESPECÍFICOS

Investigar nuevos métodos para la normalización de conceptos biomédicos con un enfoque de cross lingual.

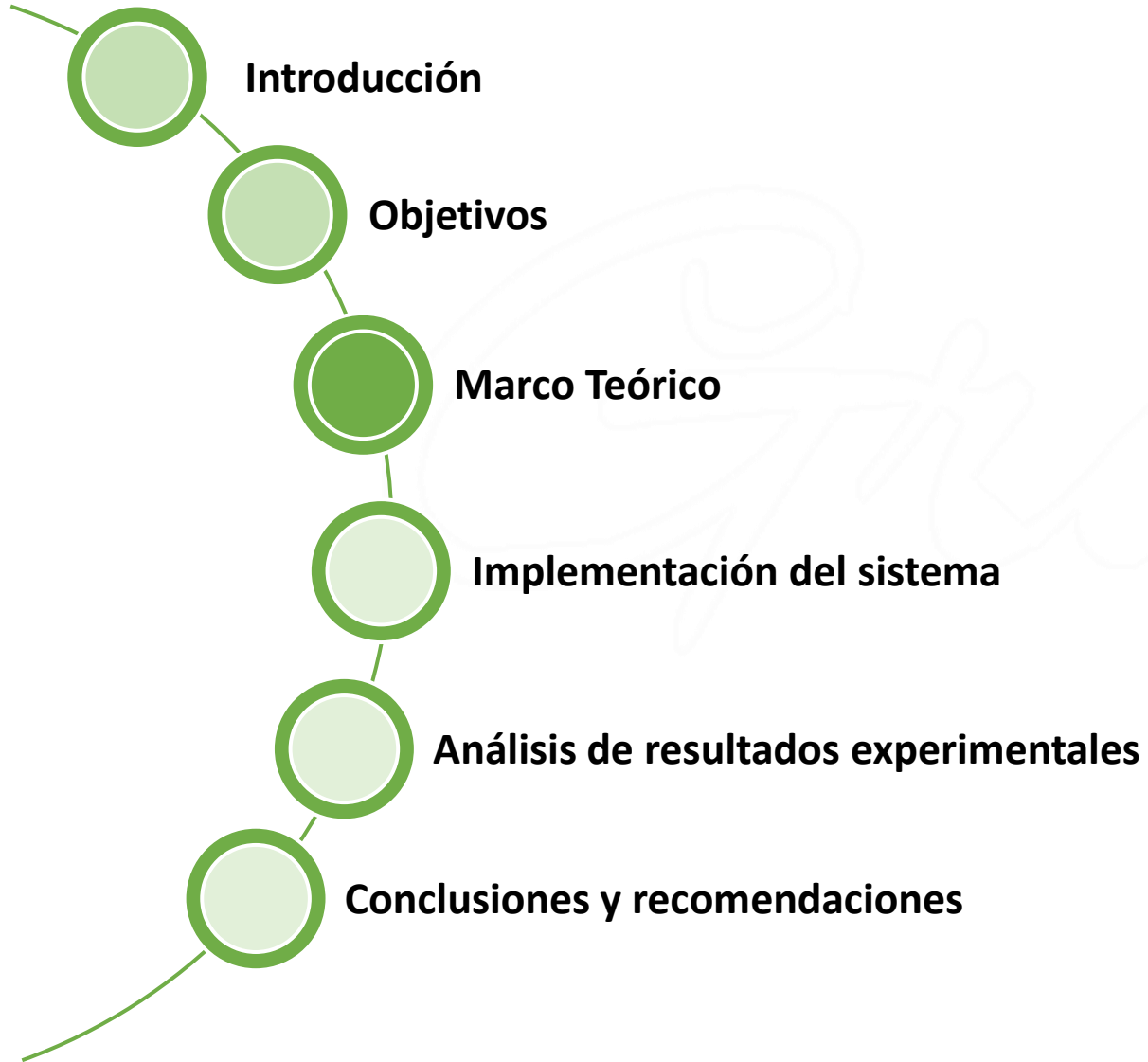
Desarrollar bases teórico conceptuales de cómo aplicar nuevos métodos para la normalización de entidades biomédicas mediante el uso de técnicas de procesamiento del lenguaje natural con un enfoque de cross-lingual.

Describir el desarrollo del sistema software frameworks y arquitecturas actuales.

Realizar la validación de los datos, la búsqueda semántica y la similitud que tuvo entre la base de datos y la petición de entrada.



CONTENIDO



UMLS (Unified Medical Language System)

UMLS tiene la característica principal de proporcionar a los desarrolladores información sobre conceptos pertenecientes al campo de la medicina, así como funcionalidades de búsqueda proporcionados para usuarios poco técnicos en el campo (National Library of Medicine, 2016).

Las tres fuentes de conocimiento UMLS las cuales se definen como:

- El Metathesaurus: Es el encargado de proporcionar una biblioteca con más de 5 millones de conceptos, términos o nombres dentro del campo de la biomedicina.
- Red Semántica: Dentro de la Red semántica del UMLS existen 133 categorías y 54 relaciones entre las categorías para etiquetar los dominós en el campo de la biomedicina.
- Herramientas Léxicas: proporciona información léxica que es utilizada para el procesamiento del lenguaje.



Átomos

Dentro del Metathesaurus existe un identificador único conocido como átomo que la National Library of Medicine (National Library of Medicine, 2016) define como “nombres conceptuales o las cadenas de cada uno de los vocabularios de origen”

AUI es un identificador único capaz de especificar y diferenciar cada concepto del uno con el otro dando así una asignación única a cada concepto, dentro del formato AUI contiene una combinación de letras números alternativos para diferenciación de los conceptos dentro de la base de datos UMLS.

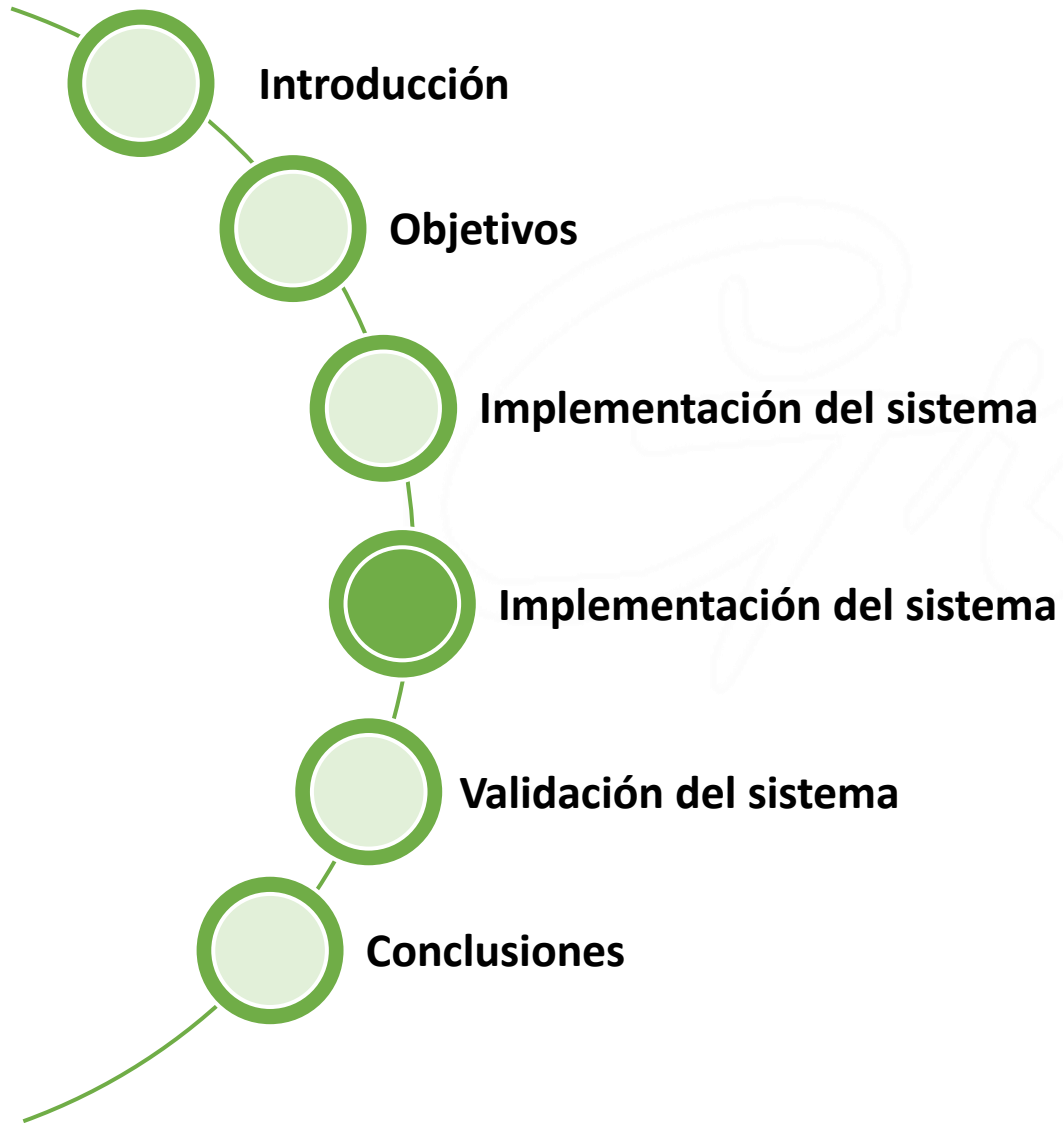


Búsqueda semántica para normalizar entidades biomédicas

La búsqueda semántica es una técnica avanzada para normalizar y categorizar términos biomédicos como genes, proteínas, enfermedades y medicamentos. La normalización es necesaria para asegurar la uniformidad y exactitud al analizar información biomédica. La búsqueda semántica utiliza modelos de lenguaje avanzados, como las redes neuronales basadas en arquitectura Transformer, eficientes para comprender las complicadas y contextuales relaciones entre los términos biomédicos. Estas técnicas utilizan un aprendizaje profundo y enfoque para comprender el sentido de las palabras en su contexto.

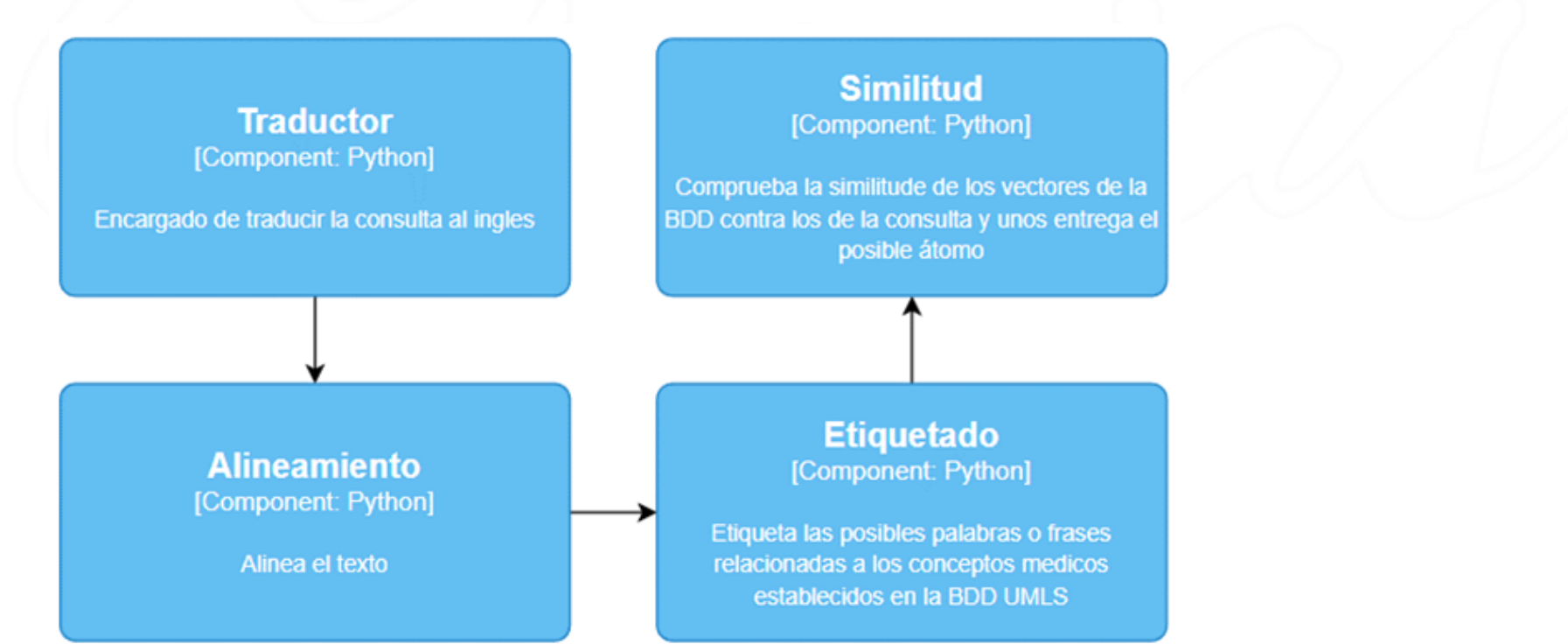


Contenido



IMPLEMENTACIÓN DEL SISTEMA

El funcionamiento del sistema recibirá un texto candidato con relación a un contexto médico. A partir de este proceso, se generará una salida que comprenderá cuatro aspectos principales:



IMPLEMENTACIÓN DEL SISTEMA

Metodología de desarrollo Ágil

La metodología Scrum es una práctica dentro del desarrollo ágil donde nos permitirá realizar actividades y tareas a lo largo el proyecto lo que nos da acceso a implementar un sistema software ágil dentro de un espacio de tiempo determinado.

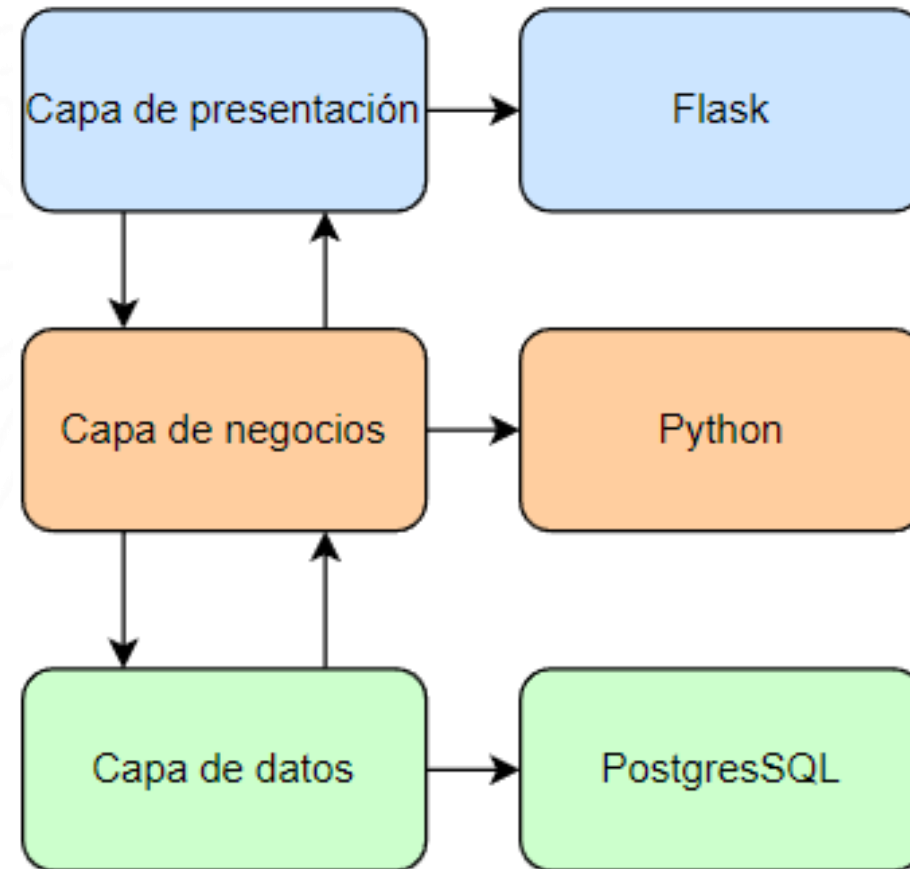


Tomado, Drew, 2023, Ventajas y desventajas de la metodología Scrum . Wearedrew.co. <https://blog.wearedrew.co/>



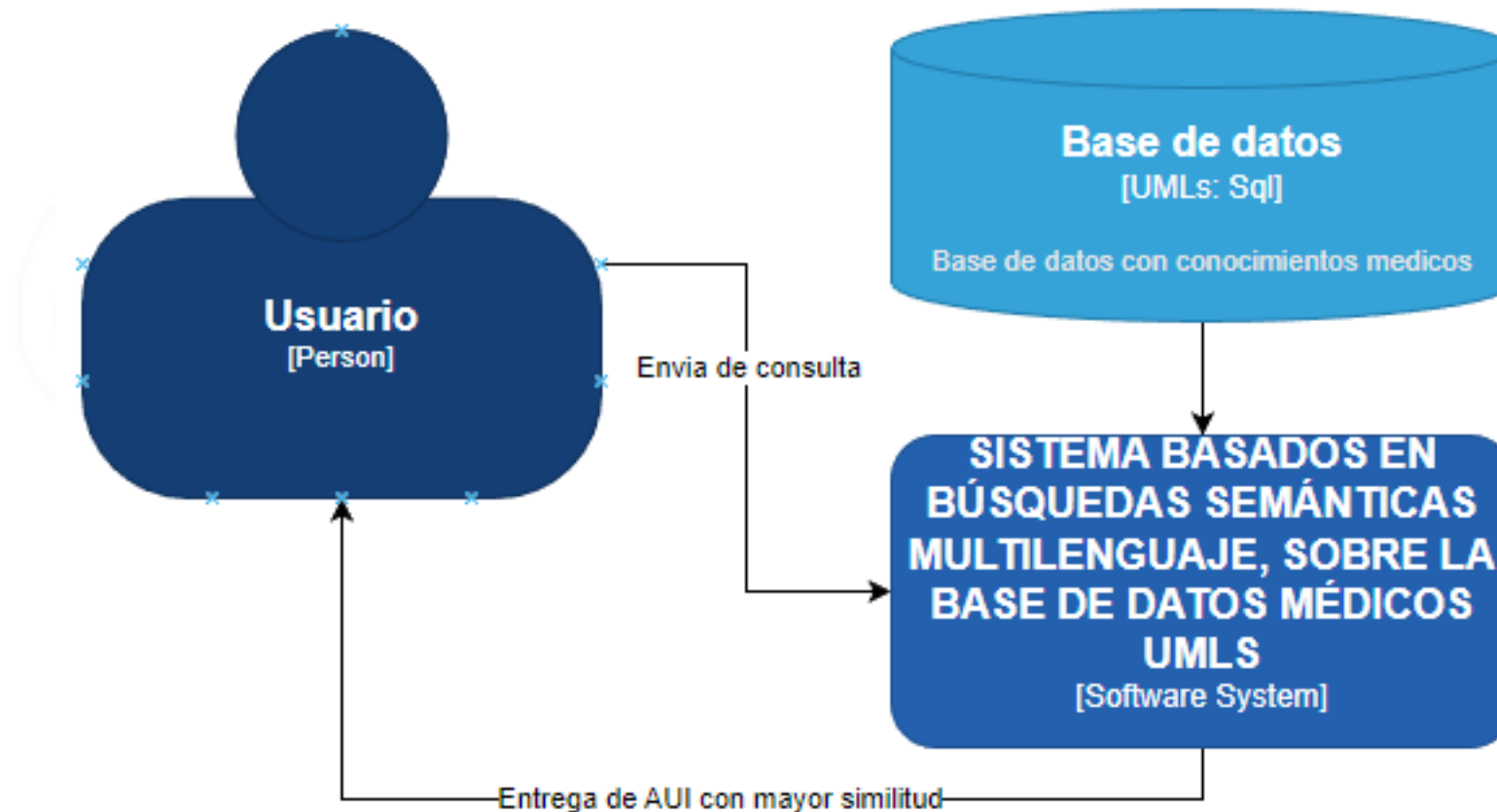
IMPLEMENTACIÓN DEL SISTEMA

ARQUITECTURA DEL SISTEMA



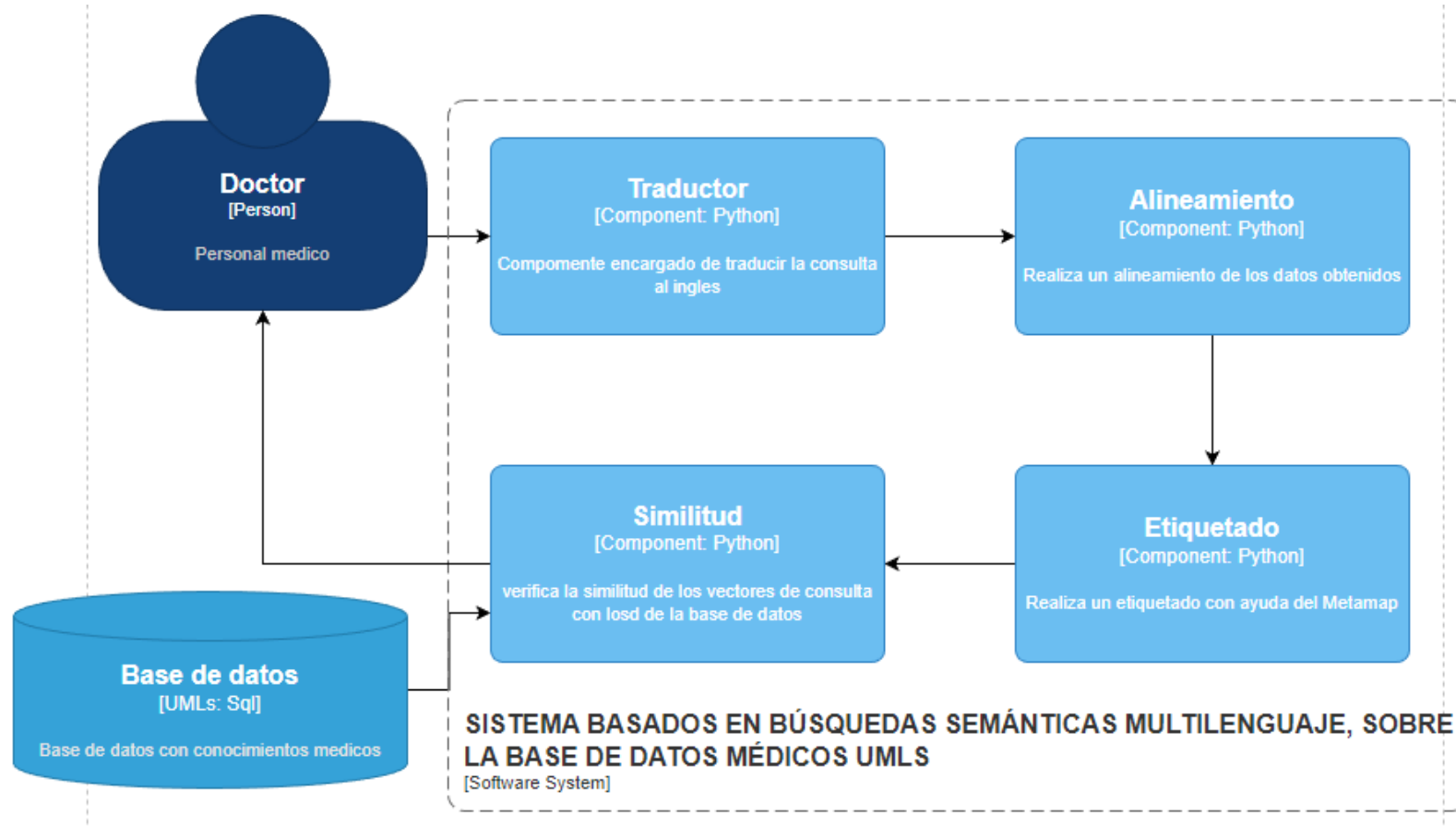
IMPLEMENTACIÓN DEL SISTEMA

MODELO C4 – NIVEL 1



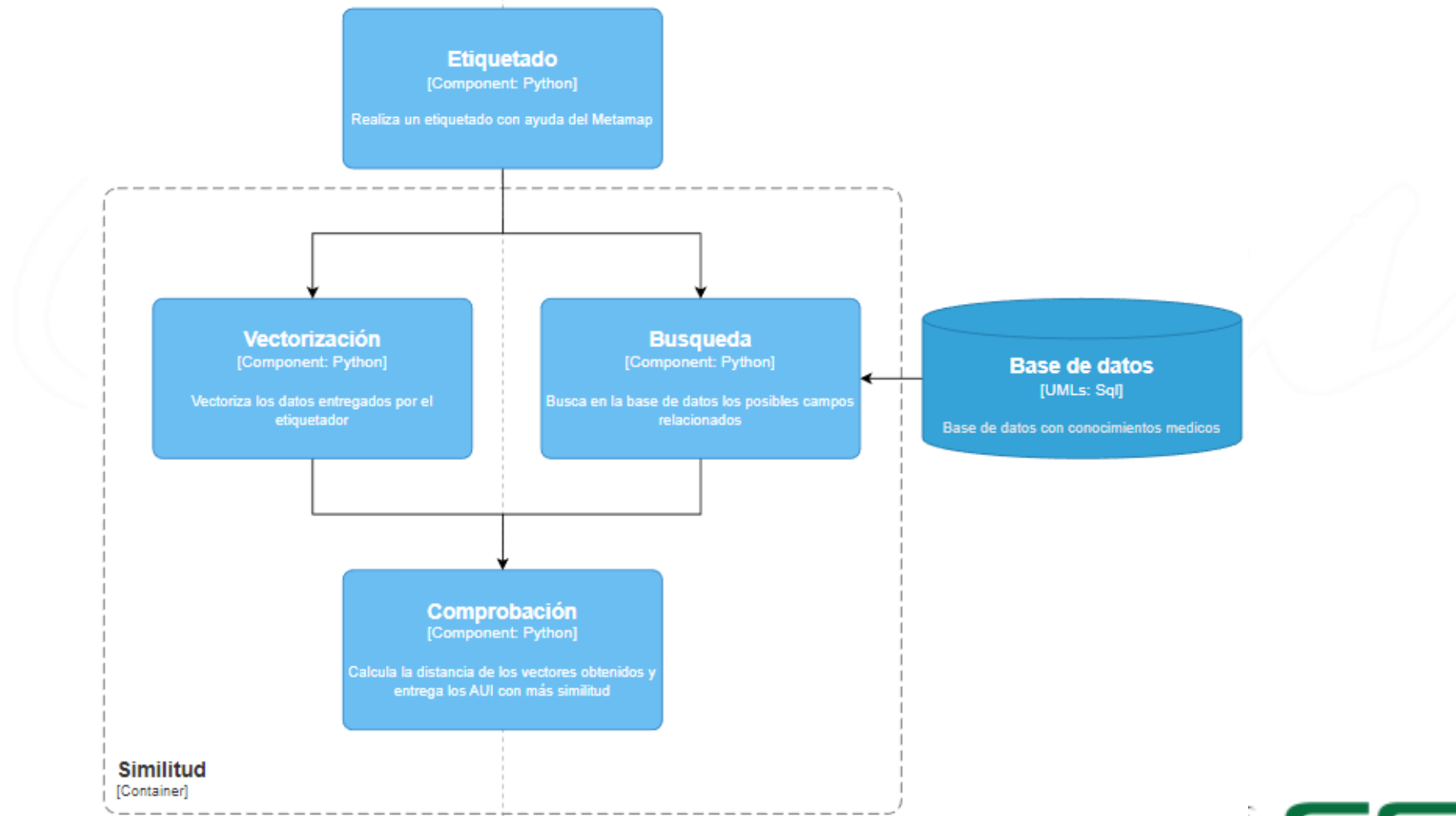
IMPLEMENTACIÓN DEL SISTEMA

MODELO C4 – NIVEL 2

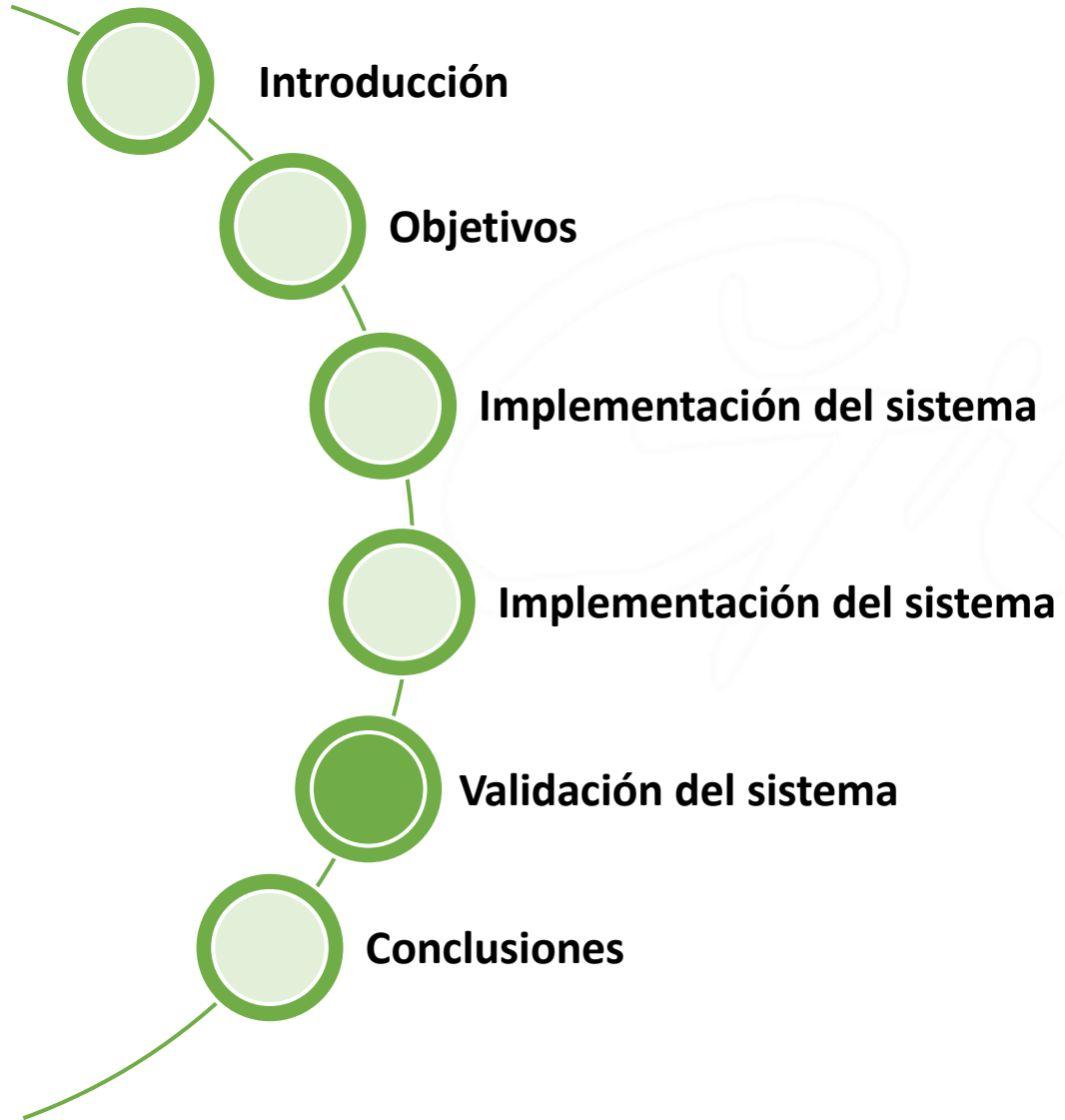


IMPLEMENTACIÓN DEL SISTEMA

MODELO C4 – NIVEL 3



Contenido



CORPUS USADO EN EL SISTEMA

La validación del sistema se apoya en un corpus que se encuentra almacenado en el repositorio de GitHub. Este corpus fue creado de manera manual, lo que significa que cada dato fue revisado y etiquetado por un individuo con conocimientos en el campo médico.



VALIDACIÓN DEL SISTEMA

METRICAS

Metrica	Obtencion del AUI
Fiabilidad	1 - 0

Métrica	0s -30s	30s - 90s	90s+
Eficiencia	1	0,5	0



VALIDACIÓN DEL SISTEMA

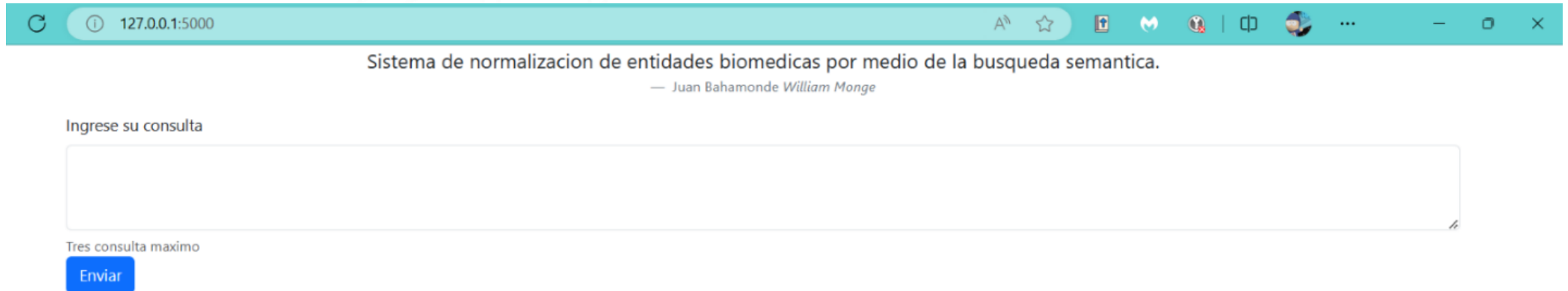
ANÁLISIS DE LOS RESULTADOS

	Fiabilidad	Eficiencia		
		Vectorizar	Búsqueda	Comprobación
Total	1	0,8	0,171875	0,828125
Total		0,6		



VALIDACIÓN DEL SISTEMA

VISUALIZACIÓN DEL SISTEMA



The screenshot shows a web browser window with the address bar displaying "127.0.0.1:5000". The page content includes a title "Sistema de normalizacion de entidades biomedicas por medio de la busqueda semantica." and a subtitle "— Juan Bahamonde William Monge". Below this is a text input field with the placeholder "Ingrese su consulta" and a blue "Enviar" button. A note below the button states "Tres consulta maximo".

127.0.0.1:5000

Sistema de normalizacion de entidades biomedicas por medio de la busqueda semantica.
— Juan Bahamonde William Monge

Ingrese su consulta

Tres consulta maximo

Enviar



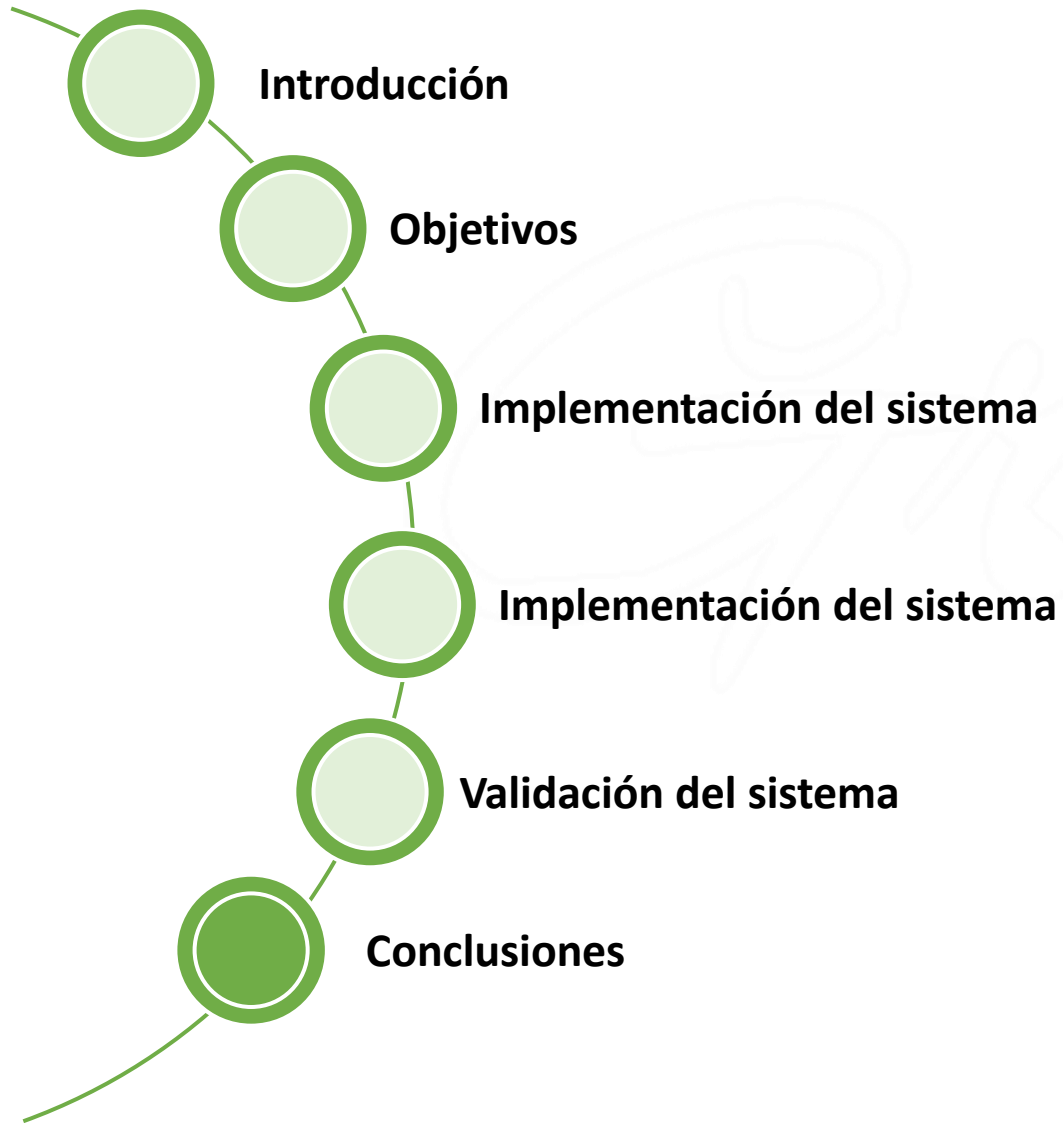
VISUALIZACIÓN DEL SISTEMA

Consulta									
[El paciente tiene dañada la estructura del miocardio]									
Traducción			Etiquetado						
[The patient has damaged the structure of myocardial]			#	CUI	Etiquetado	Frase	Inicio	Final	Puntaje
			1	C0030705	the patient	El paciente	0	11	-1000
			1	C0010957	damaged	dañada	16	23	-966
			1	C0027061	the structure of myocardial	la estructura del miocardio	24	51	-783

Átomo						
CUI	AUI	ALINEAMIENTO	TEXTO	CORRECTO	UMLS	ID DOCUMENT
C0030705	A0014128	the patient	El paciente	the	Histamine	0
C0010957	A10731772	damaged	dañada	damaged	Naglszyme	16
C0027061	A0014122	the structure of myocardial	la estructura del miocardio	of	Histamine	24



Contenido



CONCLUSIONES

Se ha logrado el objetivo de desarrollar un sistema informático para la normalización de entidades biomédicas a través de búsquedas semánticas multilenguaje, utilizando como base de datos médica el UMLS.

La construcción del marco teórico ha posibilitado una revisión detallada de conceptos sobre búsquedas semánticas y la aplicación de modelos Transformer, así como la identificación de herramientas esenciales para el desarrollo del sistema.

El diseño de la interfaz de usuario es sencillo y amigable, lo que facilita su uso por parte de los usuarios. Sin embargo, se reconocen áreas de mejora que se detallan en la sección de recomendaciones.

El enfoque de modelado C4 ha contribuido a una comprensión y visualización claras del funcionamiento del sistema, resultando en una explicación accesible para personas externas al proyecto.



¡Gracias!



ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA