



everis

# Clasificador CIE10

everis

## Introducción

- La mayor parte de la **documentación clínica** generada a pacientes en cualquier hospital es generada como **texto libre**, esto implica millones de documentos clínicos **escritos por diferentes personas, diferentes formas de expresarse y diferentes idiomas**.
- De cara a **mejorar la calidad del servicio**, las organizaciones de salud tienen la **necesidad de comparar** el servicio ofrecido entre diferentes hospitales a su cargo, así como entre diferentes organizaciones y países.
- Para cubrir esta necesidad se hace uso de **clasificaciones internacionales de enfermedades** como **CIE-10**, trasladando la información clínica a **códigos estandarizados** y por tanto comparables entre diferentes hospitales, años y organizaciones.
- En la actualidad, el proceso de codificación requiere **gran cantidad de personas** encargadas de **leer la documentación clínica** y asignarle a cada caso, y de forma **manual**, la codificación CIE-10 que le corresponde (De entre más de **140.000 diagnósticos y procedimientos** posibles).

El **objetivo del piloto** aquí presentado es crear una **herramienta de ayuda a la codificación** capaz de identificar de forma, aplicando **técnicas de PLN**, la codificación que le corresponde a un documento clínico.

# Introducción

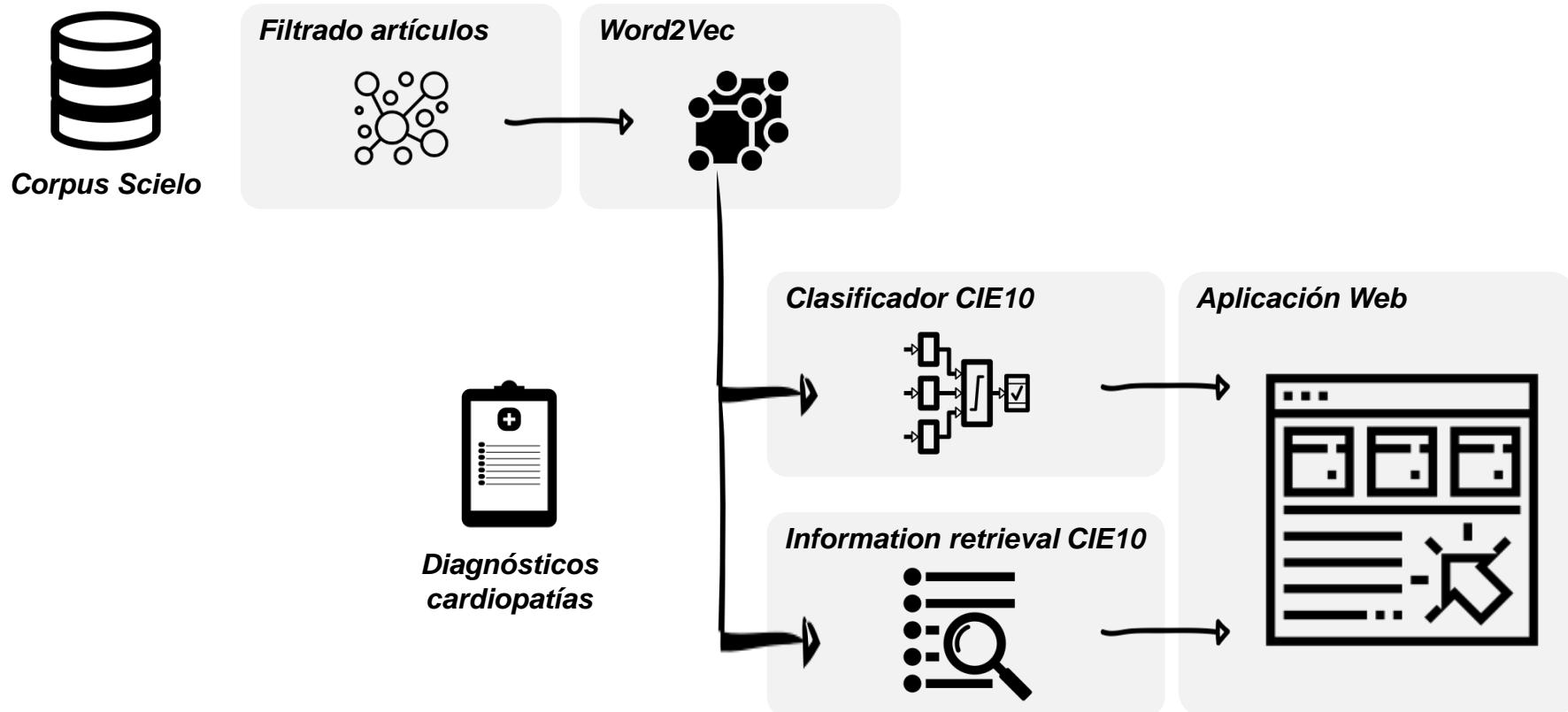
## Beneficios

Disponer de documentación clínica codificada en un estándar como **CIE-10** permite explotar la información clínica hasta niveles casi inimaginables.

Algunos **ejemplos reales** de uso son:

- Calcular las **estancias medias** asociadas a episodios clínicos (Haciendo uso de las fecha de ingreso y alta de los pacientes), comparando así entre hospitales y observando cuento tiempo dedica cada hospital a una enfermedad concreta.
- **Probabilidad de reingresos**: Dada una enfermedad concreta, qué probabilidad de reingreso existe en un hospital en los siguientes X días.
- **Gestión de gastos**: Cuántos estudios de radiología y laboratorio se han hecho a los pacientes ingresados por la enfermedad X durante el último año.
- **Estudios epidemiológicos**: En qué hospitales se ha dado más una enfermedad X durante el último periodo.

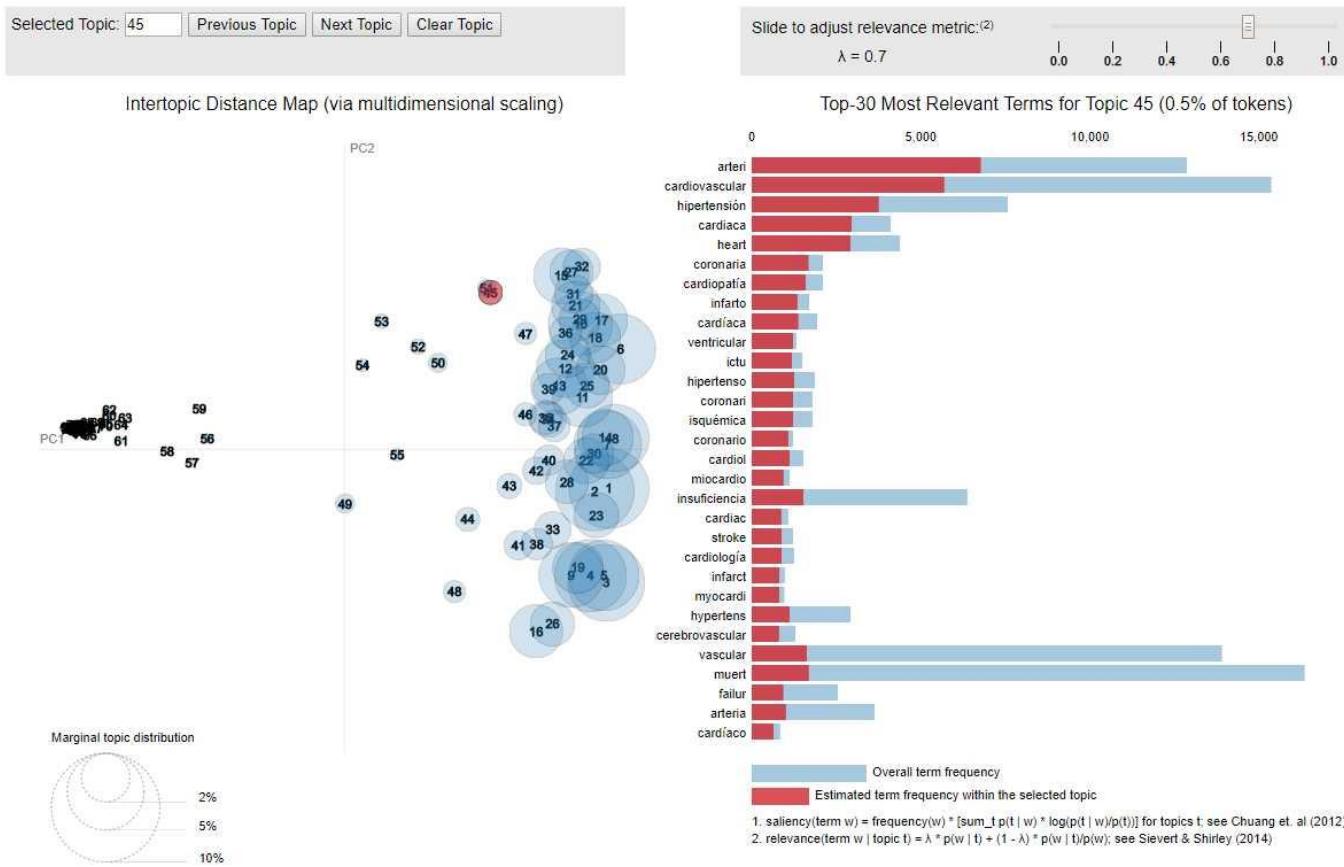
## Diseño general de la solución

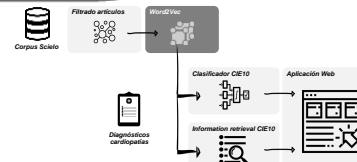


# Diseño general de la solución

## Filtrado artículos

El objetivo es analizar todos los artículos de Scielo para obtener sólo aquellos relacionados con especialidades de cardiología.

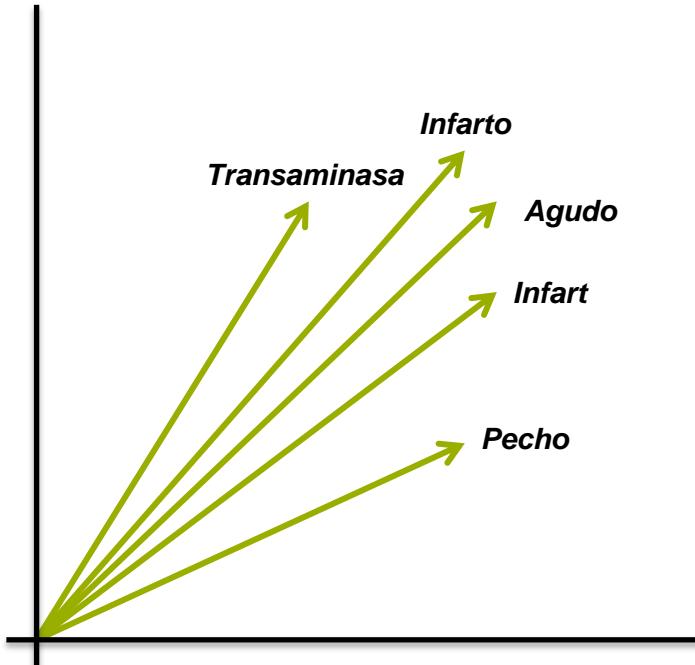




## Diseño general de la solución

### Word2vec

El objetivo es generar un **espacio de términos** con la información contenida en el corpus de Scielo **relativa a cardiología**.



Asimismo el sistema permite la consulta en función de la relación de los términos:

#### Infarto + agudo – tromboembolismo

iam - 0,828370988369  
extenso - 0,82329583168  
necrosis - 0,802765250206  
franco - 0,800782978535  
gira - 0,793510437012  
compactación - 0,793437361717

# Diseño general de la solución

## Clasificador CIE10

Para la **clasificación** en las distintas categorías se ha entrenado una **red neuronal** con la siguiente topología:



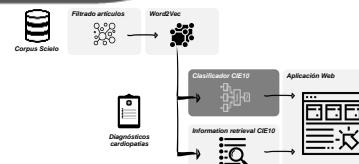
***Embedding Scielo***

***Bidirectional LSTM***

***Dropout***

***Output layer***

*Probabilidad de pertenencia a  
cada categoría*



Los resultados obtenidos por el modelo en un conjunto de test de 4.719 diagnóstico han sido:

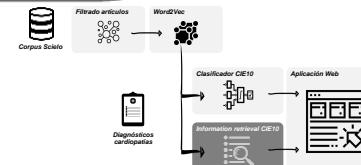
- **Aciertos: 4.189 (89%)**
- **Fallos: 530 (11%)**

La categoría que mayor error presenta es *Otras formas de enfermedad cardiaca* con una precisión del 85%.

# Diseño general de la solución

## Information Retrieval CIE10

Ante el **alto número de categorías** distintas expuestas en el CIE10 y el **bajo volumen de diagnósticos** que presentarán muchas de ellas se ha construido un **sistema de IR** que permita cubrir aquellas categorías que no sea capaz de cubrir el modelo de clasificación.



### Comparación de términos

Los términos del diagnóstico de entrada se comparan contra la descripción del CIE10 ponderando por su relevancia.

### Búsqueda de sinónimos

Mediante el word2vec de Scielo se buscan sinónimos en la descripción del CIE10.

### Similitud por LDA

Se calcula un LDA con las descripciones del CIE10 y se usa para determinar la distancia del diagnóstico a cada una de ellas.

**Embedding**

**Embedding**

**Embedding**

**Dropout**

**Dropout**

**Dropout**

**Densely connected layer**

**Dropout**

**Output layer**

*Ranking de categorías ordenadas por similitud*

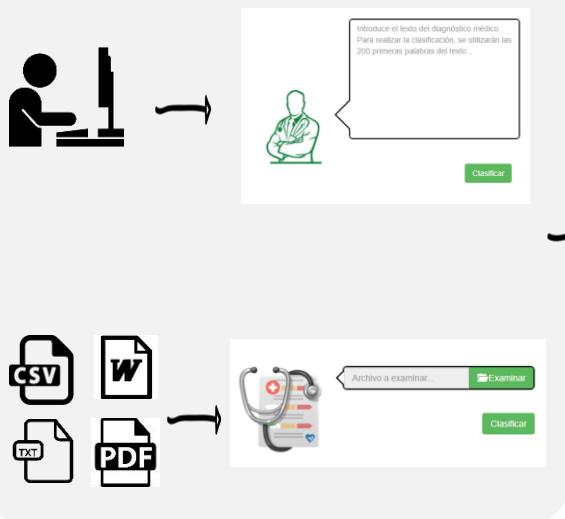


# Diseño general de la solución

## Aplicación Web

### Obtención del texto del diagnóstico

La obtención del texto del diagnóstico se puede realizar o bien introduciendo el texto de forma manual, o bien mediante el anexado de un fichero del que se realizará la extracción tomando como formatos aceptados (pdf, docx, txt y csv) en castellano o catalán



### Llamada al servicio de clasificación

Una vez obtenido el texto del diagnóstico, se invoca a un servicio web desarrollado con Python y Flask enviando como parámetro de entrada el texto a clasificar y que devolverá el resultado de la clasificación, el cual será mostrado mediante una aplicación desarrollada con HTML5, Css3, Bootstrap, AngularJs y Python

### Visualización de los datos

Se presentarán dos tablas:

- En la primera tabla se mostrará un TOP con los 3 resultados de clasificación más óptimos identificando con colores según la probabilidad de acierto en la clasificación.
- En la segunda tabla se mostrará un TOP con las 5 mejores sugerencias que nos devolverá el *Information Retrieval* de CIE-10

The screenshot shows a user interface for classification results. At the top, a message says 'El diagnóstico ha sido clasificado como:'. Below is a table with three rows:

Código CIE	Nombre	% Acierto
I80-I89	Enfermedades de vena, vasos linfáticos y nodos linfáticos, no clasificados en otra parte	87%
I10-I16	Enfermedades hipertensivas	5.5%

Below this, another message says 'El IR devuelve las siguientes sugerencias:'. Below is a table with five rows:

Código CIE	Nombre
I05-I08	Enfermedad cardíaca pulmonar y enfermedades de la circulación pulmonar
I70-I79	Enfermedades de arterias, arterias y capilares
I60-I69	Enfermedades cerebrovasculares
I80-I89	Enfermedades de vena, vasos linfáticos y nodos linfáticos, no clasificados en otra parte
I10-I16	Enfermedades hipertensivas