

# Desarrollo de un prototipo funcional de etiquetado de informes médicos con una terminología clínica reconocida mediante técnicas de PLN



**Universidad**  
Internacional  
de Valencia

13 Marzo 2023

Titulación:

Máster de Big Data y  
Ciencia de Datos

Curso académico

2022 – 2023

Alumno/a:

Medina Fernández, Fernando

Director/a de TFM: Igual Pérez,  
Román

Convocatoria:

Primera

De:

 Planeta Formación y Universidades



## Índice

Apéndice I.....	5
Apéndice II.....	21
Apéndice III.....	32
Apéndice IV .....	35
Apéndice V .....	42

# Índice de ilustraciones

Ilustración 1. Comparación de tiempos de carga de cada uno de los modelos. Elaboración propia.....	16
Ilustración 2. Comparación de tiempos de ejecución de cada uno de los modelos. Elaboración propia.....	17
Ilustración 3. Comparación de tiempos de vinculación del informe tipo mediante distancia euclidiana en cada uno de los modelos. Elaboración propia.....	18
Ilustración 4. Comparación de tiempos de vinculación del informe tipo mediante similitud coseno en cada uno de los modelos. Elaboración propia.....	19
Ilustración 5. Representación gráfica y fórmula matemática para obtener la distancia o similitud coseno. Fuente: <a href="https://www.tyrrell4innovation.ca">https://www.tyrrell4innovation.ca</a> .....	30
Ilustración 6. Fórmula matemática para obtener la distancia euclidiana o norma euclídea. Fuente: <a href="https://es.wikipedia.org">https://es.wikipedia.org</a> .....	31
Ilustración 7. Diagrama físico de la base de datos. Elaboración propia mediante MySQL Workbench.....	32
Ilustración 8. Página de inicio del prototipo. Elaboración propia.....	35
Ilustración 9. Menú de informes del prototipo. Elaboración propia. ....	36
Ilustración 10. Interfaz de creación y etiquetado de informes del prototipo. Elaboración propia. ....	36
Ilustración 11. Resultado del etiquetado de un nuevo informe. Elaboración propia.....	37
Ilustración 12. Interfaz de consulta de informes del prototipo. Elaboración propia. ....	37
Ilustración 13. Resultado de la consulta de informes. Elaboración propia.....	38
Ilustración 14. Menú de terminología clínica del prototipo. Elaboración propia.....	38
Ilustración 15. Interfaz de creación de términos clínicos del prototipo. Elaboración propia. ....	39
Ilustración 16. Interfaz de consulta de términos clínicos del prototipo. Elaboración propia. ....	39
Ilustración 17. Interfaz de visualización de términos clínicos del prototipo. Elaboración propia. ....	40
Ilustración 18. Menú de “acerca de” del prototipo. Elaboración propia. ....	40

# Índice de tablas

<i>Tabla 1. Resultados de etiquetas vinculadas para el informe tipo. Elaboración propia mediante el interfaz de usuario del prototipo.....</i>	<i>11</i>
<i>Tabla 2. Resultados de tomas de tiempos para los modelos de entrenamiento con textos clínicos a nivel de palabras completas. Elaboración propia. ....</i>	<i>12</i>
<i>Tabla 3. Resultados de umbrales y términos vinculados para los modelos de entrenamiento con textos clínicos a nivel de palabras completas. Elaboración propia.</i>	<i>12</i>
<i>Tabla 4. Resultados de tomas de tiempos para los modelos de entrenamiento con textos clínicos a nivel de subpalabras. Elaboración propia. ....</i>	<i>13</i>
<i>Tabla 5. Resultados de valores de umbras y términos vinculados para los modelos de entrenamiento con textos clínicos a nivel de subpalabras. Elaboración propia. ....</i>	<i>14</i>
<i>Tabla 6. Resultados de tomas de tiempos para los modelos de entrenamiento con textos biomédicos a nivel de palabras completas. Elaboración propia. ....</i>	<i>14</i>
<i>Tabla 7. Resultados de valores de umbrales y número de términos vinculados para los modelos de entrenamiento con textos biomédicos a nivel de palabras completas. Elaboración propia.....</i>	<i>14</i>
<i>Tabla 8. Resultados de tomas de tiempos para los modelos de entrenamiento con textos biomédicos a nivel de subpalabras. Elaboración propia. ....</i>	<i>15</i>
<i>Tabla 9. Resultados de umbrales y términos vinculados para los modelos de entrenamiento con textos biomédicos a nivel de subpalabras. Elaboración propia. ....</i>	<i>15</i>
<i>Tabla 10. Operaciones implementadas en el motor de inferencia desarrollado. Elaboración propia.....</i>	<i>22</i>
<i>Tabla 11. Funciones implementadas en el módulo de gestión de términos clínicos. Elaboración propia.....</i>	<i>23</i>
<i>Tabla 12. Funciones implementadas en el módulo de gestión de informes. Elaboración propia. ....</i>	<i>24</i>
<i>Tabla 13. Funciones implementadas en el módulo de usuarios. Elaboración propia...</i>	<i>25</i>
<i>Tabla 14. Funciones implementadas en el módulo de acceso a datos de términos clínicos. Elaboración propia. ....</i>	<i>26</i>
<i>Tabla 15. Funciones implementadas en el módulo de acceso a datos de informes. Elaboración propia.....</i>	<i>27</i>
<i>Tabla 16. Tablas, campos y tipos de datos de la base de datos. Elaboración propia..</i>	<i>34</i>
<i>Tabla 17. Tabla con el resultado al validar los datos etiquetados por el prototipo en relación con DISTEMIST. Elaboración propia. ....</i>	<i>66</i>
<i>Tabla 18. Tabla con el resultado al validar los datos etiquetados por el prototipo en relación con DISTEMIST. Elaboración propia. ....</i>	<i>90</i>

# Apéndice I

## Cuadernos de Google Colab

Se ha llevado a cabo el desarrollo y pruebas de los diferentes algoritmos y estrategias encontrados en la revisión del estado del arte para llevar a cabo la vinculación de los elementos de la terminología a los informes médicos, utilizando Google Colab (<https://colab.research.google.com/>) y los cuadernos Jupyter (<https://jupyter.org/>).

Se presentan a continuación los cuadernos generados y trabajados en el orden cronológico de su implementación:

- **Bioembeddings\_spa.ipynb:**

Cuaderno en el que se ha experimentado con modelos de obtención de vectores densos existentes (<https://github.com/PlanTL-GOB-ES/Biomedical-Word-Embeddings-for-Spanish>), con la librería *Gensim*, para comprobar cómo cargar los modelos, cómo calcular la similitud o distancia entre vectores densos, y comprobar si es posible encontrar similitudes entre palabras y grupos de palabras entre sí.

Después de analizar los resultados obtenidos en la ejecución de este cuaderno, se determina que la obtención de vectores densos mediante estos modelos, y el cálculo de similitud entre los vectores densos de grupos de palabras con significado parecido o igual (en el ámbito clínico en español), puede servir con el propósito de experimentación para encontrar grupos de palabras similares entre sí.

- **Read SnomedCT Terms.ipynb:**

Cuaderno en el que se ha experimentado con el modelo de obtención de vectores densos, con la carga de una terminología clínica específica SNOMED-CT, con la similitud coseno entre los textos, y con el almacenamiento en base de datos de los resultados.

Después de analizar los resultados obtenidos en la ejecución de este cuaderno, se determina que la obtención de vectores densos mediante estos modelos, y el cálculo de similitud coseno entre los vectores densos de terminologías clínicas y de grupos de palabras un informe médico, con significado parecido o igual (en el ámbito clínico en español), puede servir con el propósito de experimentar en la localización de grupos de palabras similares entre sí. Aquí se constata también, que es necesario ajustar adecuadamente el umbral de valor de la similitud que se tiene en cuenta, para evitar que se relacionen términos con palabras de texto que realmente no tienen que ver. A priori, una similitud coseno mayor a 0.8 puede dar resultados aceptables (aunque se requieren más pruebas).

- **LearnSpacy.ipynb:**

Cuaderno obtenido de los recursos de la librería *spacy* (<https://spacy.io/>), en el que se ha experimentado con la librería *spacy*, y con sus utilidades, como la tokenización de textos, la catalogación de palabras, el reconocimiento de entidades nombradas, la búsqueda de patrones en el texto y el uso de modelos de *transformers* en *spacy*.

- **Ejemplo\_NLP\_Radiología.ipynb:**

Cuaderno obtenido de fuentes externas, en el que se ha experimentado con la librería *spacy*, para tokenizar, eliminar las palabras irrelevantes (*stopwords*), lematizar, eliminar tildes, signos de puntuación, catalogar los textos, trabajar con *TF-IDF*, y con *BOW* (*Bag-of-Words*), trabajar con n-gramas, modelos *Word2vec* para obtener vectores densos, con *BERT Transformers* para predecir palabras omitidas en un texto (*masking*).

- **BERT Word Embeddings v2.ipynb:**

Cuaderno obtenido de fuentes externas, en el que se ha experimentado el uso de *BERT* para obtener vectores densos (*embeddings*) (McCormick, 2019).

- **TFM\_Utils.ipynb:**

Cuaderno desarrollado en el que se ha experimentado con el preprocesado del texto utilizando *spacy*, la carga y uso de los modelos *FastText Embeddings*, el modelo *Word2Vec* generalista, el modelo *BERT*, la carga de terminologías clínicas a partir de un fichero, la obtención de n-gramas, y la realización de un algoritmo para procesar los informes y obtener los términos clínicos relacionados.

Después de analizar los resultados obtenidos en la ejecución de este cuaderno, se determina cómo debe hacerse el preprocesado del texto previo y se obtiene un primer algoritmo para obtener los términos clínicos relacionados de un informe médico.

El preprocesado consiste en la eliminación de palabras irrelevantes, lematización de palabras, eliminación de tildes y de signos de puntuación.

El algoritmo consiste en realizar los siguientes pasos:

1. Preprocesado del cuerpo del informe,
2. División en n-gramas o grupos de palabras,
3. Obtención de vectores densos de los n-gramas,
4. Comparación de estos con los vectores densos de los términos del vocabulario,
5. Selección de los términos identificados con una similitud coseno mayor a 0.9,
6. Eliminación de duplicados posterior.

Se experimenta además con otros pasos adicionales en la obtención de los términos vinculados, como el purgado de términos clínicos obtenidos similares entre sí (manteniendo el término con mayor similitud), para reducir el número de términos clínicos vinculados que se obtienen como resultado. No obstante, este post procesado

debe analizarse en mayor detalle porque tiene el riesgo de descartar términos clínicos que pudieran ser relevantes realmente.

En este punto **se empiezan a detectar limitaciones** en la aproximación. La vinculación funciona bien para encontrar palabras similares en el texto y en el vocabulario para proceder al etiquetado.

Al ejecutar el algoritmo de vinculación se vinculan al texto etiquetas similares en contenido desde el punto de vista morfológico pero que no se corresponden realmente, por ejemplo, por ser de partes anatómicas diferentes del cuerpo. No es lo mismo un “carcinoma en el hígado” que un “carcinoma en la mama”, sin embargo, ante la palabra “carcinoma” en el texto se devuelven ambas etiquetas erróneamente. Esto se acentúa cuando se compara la similitud de palabras atómicas con palabras del vocabulario.

Por otro lado, a la hora de obtener la representación densa de los grupos de palabras (n-gramas) se podrían procesar conjuntamente como un vector denso que representa una palabra compuesta (por ejemplo, vector de “abdomen distendido”) o bien como un vector que representa la media de los vectores densos de las palabras aisladas (media de vector de “abdomen” y vector de “distendido”). El caso más adecuado es el segundo.

- **TFM\_Utils\_BERT.ipynb:**

Cuaderno desarrollado en el que se ha experimentado con el preprocesado del texto utilizando spacy, la carga y uso del modelo *BERT*, y alguna prueba del modelo *BETO-Galen* adaptado al español en el ámbito clínico (<https://github.com/guilopgar/ClinicalCodingTransformerES>), la carga de terminologías clínicas a partir de un fichero, la obtención de n-gramas, y la realización de un algoritmo para procesar los informes y obtener los términos clínicos relacionados.

- **Feature Embedding using BETO-Galén in TensorFlow.ipynb:**

Cuaderno desarrollado a partir de otro de partida enfocado en *BERT* (<https://medium.com/@aieeshashafique/feature-extraction-from-bert-25887ed2152a>) en el que se ha experimentado con el preprocesado del texto utilizando spacy, la carga y uso del modelo *BETO-Galen* adaptado al español en el ámbito clínico (para la obtención de los vectores densos), y la realización de un algoritmo para procesar los informes y obtener los términos clínicos relacionados.

Después de analizar los resultados obtenidos en la ejecución de este cuaderno, se determina cómo debe utilizarse el modelo *BETO-Galén* para obtener los vectores densos o *embeddings* de una palabra o grupo de palabras. El procedimiento es cargar el modelo, utilizar el modelo cargado para predecir a partir del texto de partida tokenizado (identificadores), junto con una lista de “0” para identificar cuántas palabras hay en la misma frase. Esto da como resultado una lista de vectores densos, una por

cada palabra de entrada, de la que se obtiene la media de dichos vectores densos para que sea el vector denso resultado.

Adicionalmente, se prueban los resultados utilizando dicho modelo con el catálogo de terminología clínica utilizada y el algoritmo para obtener la vinculación de términos clínicos. El resultado obtenido es bastante pobre puesto que se relacionan muchos términos clínicos que no tienen relación real con lo indicado en el informe y no es posible establecer ningún umbral válido para mejorar este resultado, en lo que a la similitud coseno se refiere.

- **TFM\_Benchmarking\_Spacy\_Models v1.ipynb:**

Cuaderno desarrollado para comprobar los resultados y los tiempos de ejecución de los diferentes modelos disponibles de la librería *spacy*, en español. El preprocesado del texto se realiza con la librería *spacy*, con tres modelos disponibles para el preprocesado y división en frases de los textos (<https://spacy.io/models/es>).

Después de analizar los resultados obtenidos en la ejecución de este cuaderno, se determina que el tiempo de carga y ejecución de los tres modelos es similar, y que el modelo más extenso o *large* “*es\_core\_news\_lg*” proporciona un mejor preprocesado del texto, por lo que se determina que este es el más adecuado para su uso en el prototipo (algunas palabras como grasa, páncreas, enteritis, ascitis... se preprocesan correctamente, no así en los modelos *small* y *medium*).

- **Fasttext\_Subwords\_Test.ipynb:**

Cuaderno desarrollado para experimentar con la carga del modelo, y el cálculo de vectores densos de los modelos basados en cálculo de vectores densos a partir de sub-palabras con codificación *Byte-Pair Encoding* (*BPE subwords*) (<https://github.com/PlanTL-GOB-ES/Biomedical-Word-Embeddings-for-Spanish>).

- **TFM\_Benchmarking\_Fasttext** **v2.ipynb,**  
**TFM\_Benchmarking\_Fasttext\_subwords** **v1.ipynb,**  
**TFM\_Benchmarking\_Fasttext\_BIO** **v1.ipynb** **y**  
**TFM\_Benchmarking\_Fasttext\_BIO\_subwords v1.ipynb:**

Cuadernos desarrollados para comprobar los resultados y los tiempos de los diferentes modelos disponibles de obtención de vectores densos, en el ámbito clínico/biomédico en español. En unos casos, se utilizan los modelos que trabajan a nivel de palabras completas (*Word embedding*), y en otros casos, de los modelos elegidos se toman los modelos basados en cálculo de vectores densos a partir de subpalabras con codificación *Byte-Pair Encoding* (*BPE subwords*).

El preprocesado del texto se realiza con la librería *spacy*, con el modelo más extenso disponible para el preprocesado y división en frases de los textos (<https://spacy.io/models/es>).



Se cargan los catálogos de terminologías clínicas indicadas (problemas de salud en el ámbito hospitalario) con diferentes modelos disponibles guardando en fichero los vectores densos de todos los términos clínicos para agilizar la ejecución tomando tiempos de cada paso del proceso.

Se cargan seis modelos con cada tipo (a nivel de palabras o de subpalabras) y datos de entrenamiento (textos biomédicos y textos clínicos) a partir de los modelos que se han entrenado mediante textos clínicos, en sus variantes de entrenamiento *CBOW* y *Skip-gram*, con generación de vectores densos de 50, 100 y 300 dimensiones.

Después se ejecuta con cada modelo el algoritmo de obtención de términos clínicos relacionados para un informe radiológico tipo con el catálogo precargado previamente. De estas operaciones se toman tiempos y se comparan los resultados con los esperados para dicho informe médico.

Se destaca que para la similitud entre palabras o grupo de palabras se utilizan dos tipos de medidas, la distancia euclidiana y la similitud coseno.

Las frases y etiquetas del informe tipo utilizado son las siguientes:

*“Hay evidencia de hipoatenuación hepática difusa compatible con **infiltración grasa** .*

*No hay dilatación de los conductos biliares intra o extrahepáticos.*

*El paciente se encuentra en estado post **colecistectomía**.*

*El bazo es normal.*

*El páncreas es de contorno y características de atenuación normales.*

*No hay evidencia de **masa suprarrenal**.*

*Hay una **hernia supraumbilical** de tamaño moderado que contiene grasa.*

*Los riñones son normales en tamaño, forma y configuración.*

*No se identifican **cálculos renales** ni **ureterales**.*

*No hay **hidrouréter** ni **hidronefrosis**.*

*No hay evidencia de **apendicitis**.*

*Hay varias asas de intestino delgado llenas de líquido, compatibles con una **enteritis** leve.*

*No hay **engrosamiento de la pared intestinal**.*

*No hay evidencia de **obstrucción del intestino delgado o grueso**.*

*No hay evidencia de **ascitis** abdominal o **linfadenopatía**.*

No hay evidencia de **masa vesical** intrínseca o extrínseca .

No hay **ascitis** pélvica ni **linfadenopatía** .

El útero y los ovarios no presentan ninguna anomalía.

Las imágenes de las bases pulmonares no muestran evidencia de **masa pleural** o **parenquimatosa** .

No hay **derrames pleurales**.

Hay **cicatrices** en el lóbulo medio derecho y en la língula, así como en ambas bases pulmonares.

Las estructuras óseas están libres de lesiones líticas o blásticas.

Se observan cambios degenerativos multinivel en la columna toracolumbar.

Se observan **calcificaciones** dispersas en la aorta y en sus principales ramas, compatibles con la **aterosclerosis**.”

Después de analizar los resultados obtenidos en la ejecución de estos cuadernos se obtienen los siguientes resultados:

En el informe tipo definido los resultados son aceptables, aunque se detectan algunas detecciones incorrectas por similitud de patologías que mencionan partes anatómicas diferentes: “calcificación de mama”, “dolor abdominal”, “engrosamiento del tallo hipofisario”, “espondilolistesis degenerativa”, “voz anormal”, entre otros. Por tanto, se deben hacer más pruebas con otros informes médicos.

Fuente preprocesada	Término	Distancia/similitud
apendicitis	apendicitis	100%
ascitis	ascitis	100%
lingular base	bronquiectasias en ambas bases	92.4848%
cambio degenerativo multinivel	cambios degenerativos en columna	90.24560000000001%
cicatriz	cicatriz	100%
calculo renal	cálculo renal	100%
derrame pleural	derrame pleural	100%
dilatacion conducto biliar	dilatación del conducto de Wirsung	90.6955%
dilatacion conducto biliar intro	dilatación, incluso quística, de los ductos biliares	90.9952%
asa intestino delgado lleno	edemas en asas de intestino delgado	90.2751%
enteritis leve	escleritis leve	93.3063%
degenerativo multinivel	espondilolistesis degenerativa	90.984%
enteritis	gastroenteritis	90.5493%
hernia supraumbilical	hernia paraumbilical	98.5847%
hidronefrosis	hidronefrosis	100%
obstruccion intestino delgado	invaginación de intestino delgado	95.757%
lesion	lesiones	100%
lesion litico	lesiones líticas	100%
linfadenopatía	linfadenopatía	100%
masa vesical	masa prostática	91.0735%
masa suprarrenal	masa suprarrenal	100%
evidenciar masa	masa testicular	93.2371%
vesical intrinseco	mullerianosis vesical	90.2918%
evidenciar obstruccion intestino delgado	obstrucción del intestino delgado	96.3749%
obstruccion intestino	obstrucción intestinal	95.4534%

*Tabla 1. Resultados de etiquetas vinculadas para el informe tipo. Elaboración propia mediante el interfaz de usuario del prototipo.*

En la tabla se aprecia cómo el modelo no se comporta correctamente en algunas de las etiquetas vinculadas a algunas partes del texto.

**Modelos que parten de entrenamiento con textos clínicos, a nivel de palabras completas:**

Entrenamiento	Dimensiones embedding	Tiempo carga (sg)	Tiempo ejecución (sg)	Tiempo vinculación – distancia euclidiana (sg)	Tiempo vinculación – similitud coseno (sg)
<b>CBOW</b>	50	15	52	7	14
<b>CBOW</b>	100	28	53	7	14
<b>CBOW</b>	300	53	52	7	14
<b>Skip-gram</b>	50	18	52	7	13
<b>Skip-gram</b>	100	25	49	7	13
<b>Skip-gram</b>	300	49	52	7	14

*Tabla 2. Resultados de tomas de tiempos para los modelos de entrenamiento con textos clínicos a nivel de palabras completas. Elaboración propia.*

**Tiempo carga (sg):** Tiempo de carga del modelo en segundos.

**Tiempo ejecución (sg):** Tiempo de obtención de los vectores densos de todos los términos clínicos del vocabulario (3600 elementos).

**Tiempo vinculación (sg):** Tiempo de obtención de los términos clínicos relacionados con el informe médico tipo. Los tiempos se miden en dos escenarios, utilizando la fórmula de cálculo de la distancia euclidiana y la similitud coseno.

entrenamiento	Dimensiones embedding	Umbral Distancia Euclidiana	Umbral Similitud Coseno	Términos vinculados (distancia/similitud)
<b>CBOW</b>	50	3.6	0.95	17/21
<b>CBOW</b>	100	3.6	0.95	11/14
<b>CBOW</b>	300	3.6	0.95	11/14
<b>Skip-gram</b>	50	1.2	0.975	17/12
<b>Skip-gram</b>	100	1.2	0.975	11/11
<b>Skip-gram</b>	300	1.2	0.975	11/11

*Tabla 3. Resultados de umbrales y términos vinculados para los modelos de entrenamiento con textos clínicos a nivel de palabras completas. Elaboración propia.*

**Umbral distancia euclidiana:** Valor de distancia máxima para vincular dos términos por cercanía semántica.

**Umbral similitud coseno:** Valor de similitud coseno mínimo para vincular dos términos por cercanía semántica.

**Términos vinculados (distancia/similitud):** Número de términos vinculados identificados en el texto con respecto a la terminología clínica en función del modo de medición de la similitud.

Como se observa, el tiempo de carga depende del tamaño del modelo, el tiempo de ejecución es similar en todos los modelos, y el tiempo de vinculación es la mitad cuando se utiliza el cálculo de distancia euclidiana con respecto al uso del cálculo de la similitud coseno.

Por otro lado, en cuanto a los umbrales, se han determinado empíricamente para tratar de encontrar unos umbrales que obtengan resultados aceptables, y comparables entre sí. Así, se observa que el modelo mejora en lo que respecta a la detección de términos vinculados, conforme se aumenta el número de dimensiones de los vectores densos. El modo de entrenamiento no afecta en la mejora de la vinculación.

**Modelos que parten de entrenamiento con textos clínicos, a nivel de subpalabras codificadas BPE:**

entrenamiento	Dimensiones embedding	Tiempo carga (sg)	Tiempo ejecución (sg)	Tiempo vinculación – distancia euclidiana (sg)	Tiempo vinculación – similitud coseno (sg)
CBOW	50	3	60	9	19
CBOW	100	8	60	10	21
CBOW	300	19	60	10	22
Skip-gram	50	4	60	9	19
Skip-gram	100	6	60	11	21
Skip-gram	300	20	60	11	22

Tabla 4. Resultados de tomas de tiempos para los modelos de entrenamiento con textos clínicos a nivel de subpalabras. Elaboración propia.

entrenamiento	Dimensiones embedding	Umbral Distancia Euclidiana	Umbral Similitud Coseno	Términos vinculados (distancia/similitud)
CBOW	50	4.5	0.93	22/15
CBOW	100	3.6	0.95	11/11
CBOW	300	3.6	0.95	10/11
Skip-gram	50	1.2	0.975	34/11
Skip-gram	100	1.2	0.975	19/10
Skip-gram	300	1.2	0.975	11/9

*Tabla 5. Resultados de valores de umbras y términos vinculados para los modelos de entrenamiento con textos clínicos a nivel de subpalabras. Elaboración propia.*

Como se observa, el tiempo de carga depende del tamaño del modelo, el tiempo de ejecución es similar en todos los modelos, y el tiempo de vinculación es la mitad cuando se utiliza el cálculo de distancia euclidiana con respecto al uso del cálculo de la similitud coseno.

Por otro lado, en cuanto a los umbrales, se han determinado empíricamente para tratar de encontrar unos umbrales que obtengan resultados aceptables, y comparables entre sí. Así, se observa que el modelo mejora en lo que respecta a la detección de términos vinculados, conforme se aumenta el número de dimensiones de los vectores densos. El modo de entrenamiento no afecta en la mejora de la vinculación.

**Modelos que parten de entrenamiento con textos biomédicos, a nivel de palabras completas:**

entrenamiento	Dimensiones embedding	Tiempo carga (sg)	Tiempo ejecución (sg)	Tiempo vinculación – distancia euclidiana (sg)	Tiempo vinculación – similitud coseno (sg)
<b>CBOW</b>	50	60	60	7	14
<b>CBOW</b>	100	120	60	7	14
<b>CBOW</b>	300	180	60	7	14
<b>Skip-gram</b>	50	60	48	9	20
<b>Skip-gram</b>	100	60	48	9	20
<b>Skip-gram</b>	300	120	60	9	20

*Tabla 6. Resultados de tomas de tiempos para los modelos de entrenamiento con textos biomédicos a nivel de palabras completas. Elaboración propia.*

entrenamiento	Dimensiones embedding	Umbral Distancia Euclidiana	Umbral Similitud Coseno	Términos vinculados (distancia/similitud)
<b>CBOW</b>	50	3.6	0.95	11/18
<b>CBOW</b>	100	3.6	0.95	11/13
<b>CBOW</b>	300	3.6	0.95	11/11
<b>Skip-gram</b>	50	1.2	0.95	+50/12
<b>Skip-gram</b>	100	1.2	0.95	17/12
<b>Skip-gram</b>	300	1.2	0.95	10/11

*Tabla 7. Resultados de valores de umbrales y número de términos vinculados para los modelos de entrenamiento con textos biomédicos a nivel de palabras completas. Elaboración propia.*

Como se observa, el tiempo de carga depende del tamaño del modelo, el tiempo de ejecución es similar en todos los modelos, y el tiempo de vinculación es la mitad cuando se utiliza el cálculo de distancia euclidiana con respecto al uso del cálculo de la similitud coseno.

Por otro lado, en cuanto a los umbrales, se han determinado empíricamente para tratar de encontrar unos umbrales que obtengan resultados aceptables, y comparables entre sí. Así, se observa que el modelo mejora en lo que respecta a la detección de términos vinculados, conforme se aumenta el número de dimensiones de los vectores densos. El modo de entrenamiento no afecta en la mejora de la vinculación.

#### Modelos que parten de entrenamiento con textos biomédicos, a nivel de subpalabras codificadas BPE:

entrenamiento	Dimensiones embedding	Tiempo carga (sg)	Tiempo ejecución (sg)	Tiempo vinculación – distancia euclidiana (sg)	Tiempo vinculación – similitud coseno (sg)
<b>CBOW</b>	50	3	56	10	19
<b>CBOW</b>	100	7	58	9	19
<b>CBOW</b>	300	21	55	10	19
<b>Skip-gram</b>	50	4	51	9	18
<b>Skip-gram</b>	100	7	52	8	19
<b>Skip-gram</b>	300	16	53	10	19

Tabla 8. Resultados de tomas de tiempos para los modelos de entrenamiento con textos biomédicos a nivel de subpalabras. Elaboración propia.

entrenamiento	Dimensiones embedding	Umbral Distancia Euclidiana	Umbral Similitud Coseno	Términos vinculados (distancia/similitud)
<b>CBOW</b>	50	10	0.85	+50/+50
<b>CBOW</b>	100	10	0.85	35/34
<b>CBOW</b>	300	10	0.85	20/27
<b>Skip-gram</b>	50	1.2	0.95	+50/22
<b>Skip-gram</b>	100	1.2	0.95	42/18
<b>Skip-gram</b>	300	1.2	0.95	19/18

Tabla 9. Resultados de umbrales y términos vinculados para los modelos de entrenamiento con textos biomédicos a nivel de subpalabras. Elaboración propia.

Como se observa, el tiempo de carga depende del tamaño del modelo, el tiempo de ejecución es similar en todos los modelos, y el tiempo de vinculación es la mitad cuando se utiliza el cálculo de distancia euclidiana con respecto al uso del cálculo de la similitud coseno.

A continuación, se presentan algunas gráficas con la información obtenida.

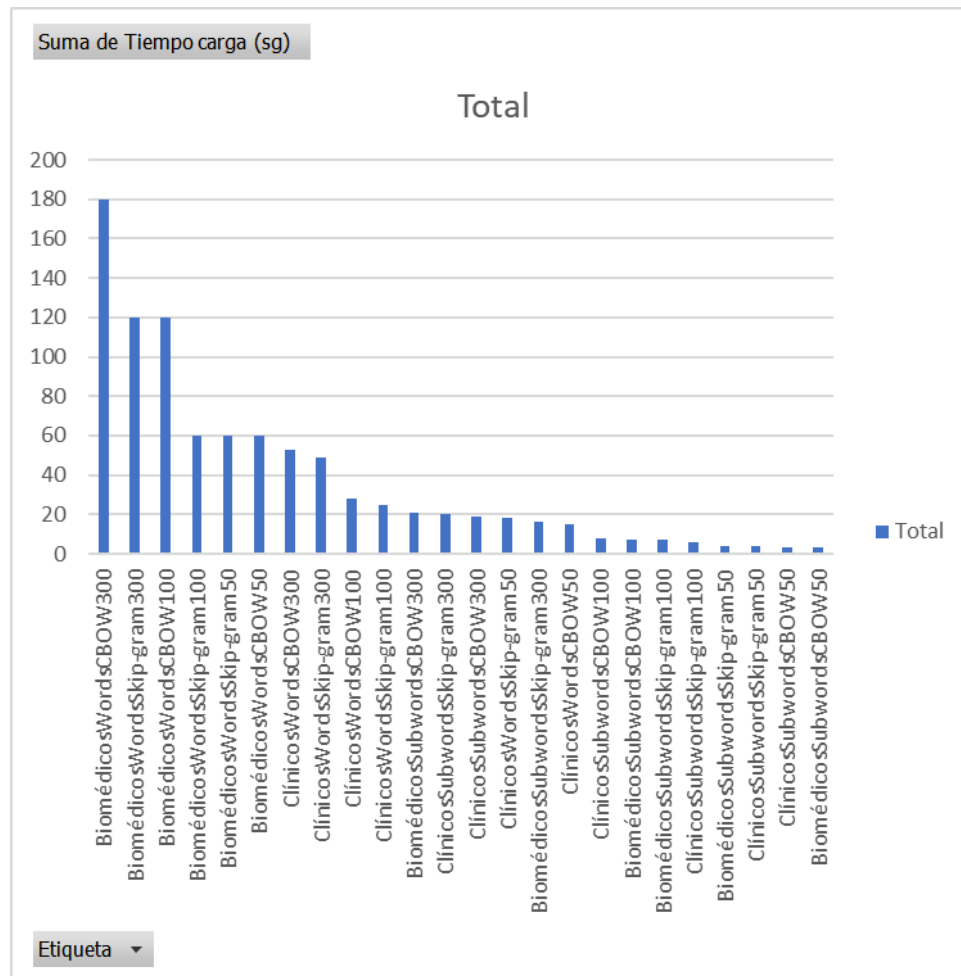


Ilustración 1. Comparación de tiempos de carga de cada uno de los modelos. Elaboración propia.

Se observa que los modelos más pesados son los que más tardan en ser cargados en memoria. En este caso los modelos entrenados con mayor cantidad de datos del ámbito biomédico y para trabajar a nivel de palabras de mayor vocabulario son los más pesados.



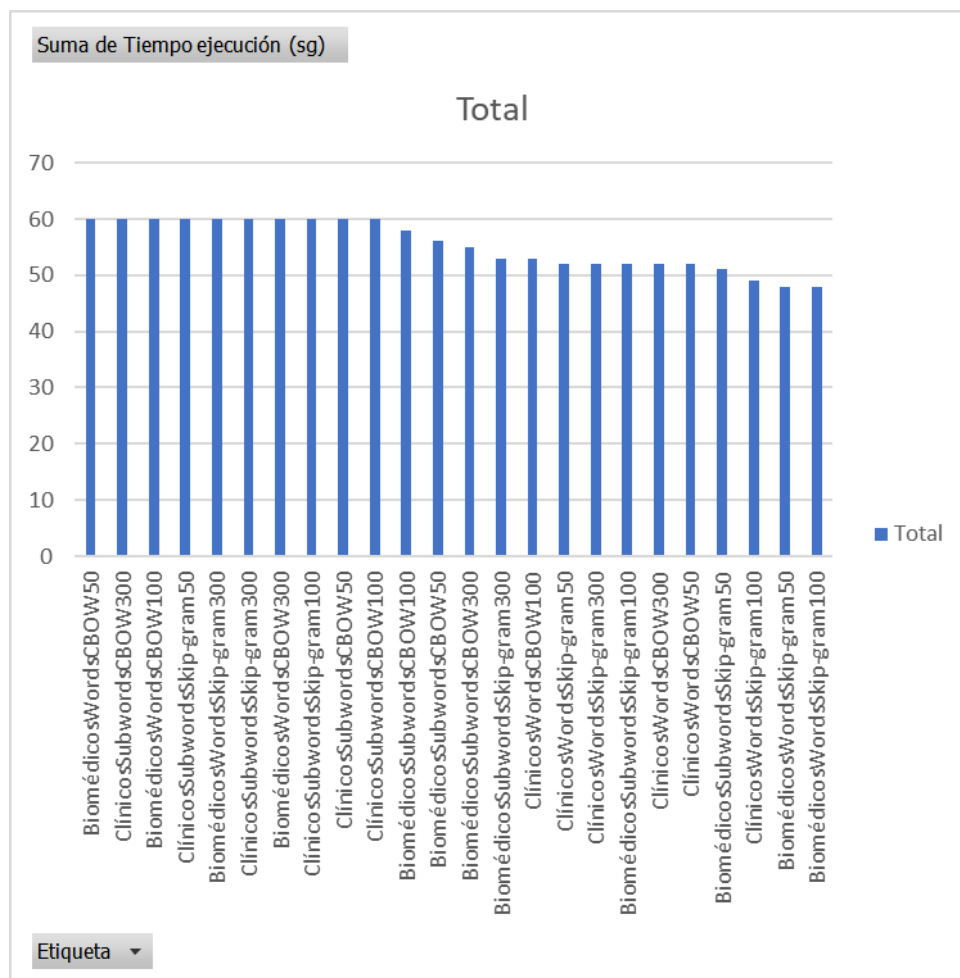
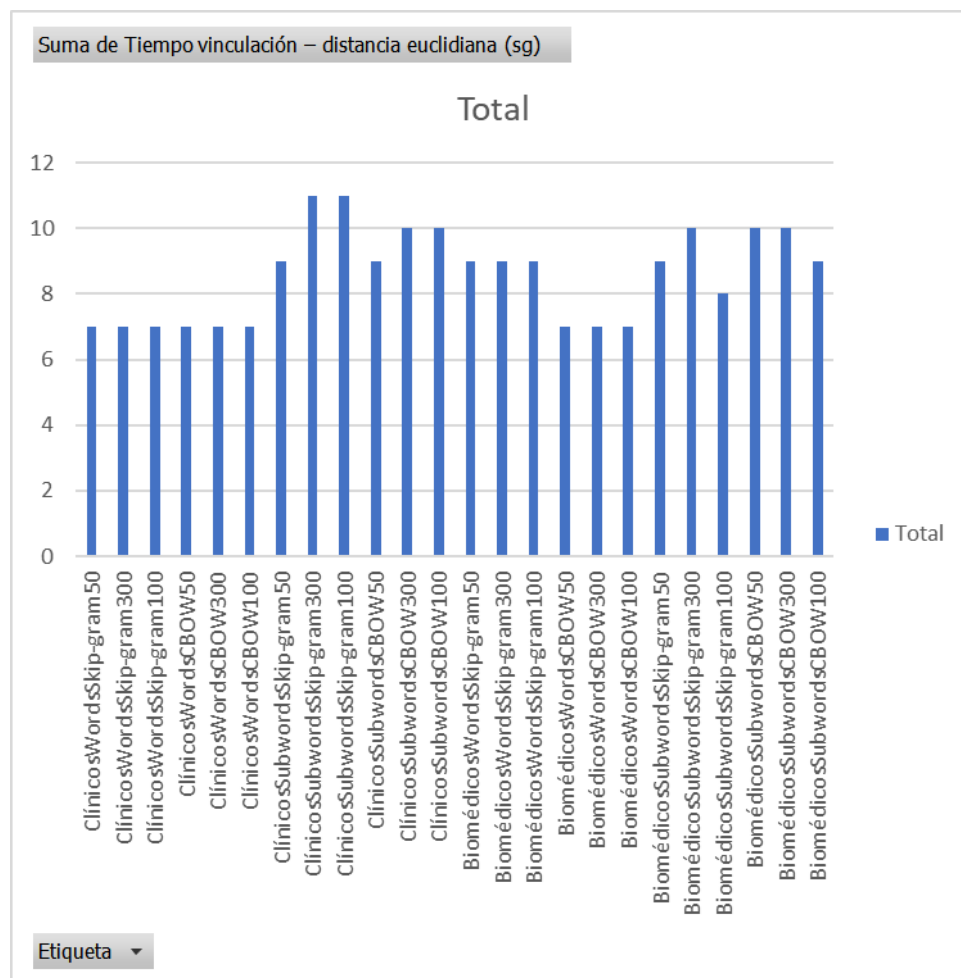


Ilustración 2. Comparación de tiempos de ejecución de cada uno de los modelos. Elaboración propia.

Se observa que el tiempo de obtención de los vectores densos del vocabulario es similar en todos los modelos, aunque los modelos con el entrenamiento *skip-gram* parecen ser los más rápidos en general.



*Ilustración 3. Comparación de tiempos de vinculación del informe tipo mediante distancia euclidiana en cada uno de los modelos. Elaboración propia.*

Se observa que el tiempo de vinculación del informe tipo utilizando la distancia euclidiana como medida es muy similar en todos los modelos, aunque los modelos los modelos que trabajar a nivel de palabras parecen ser algo más rápidos en general.

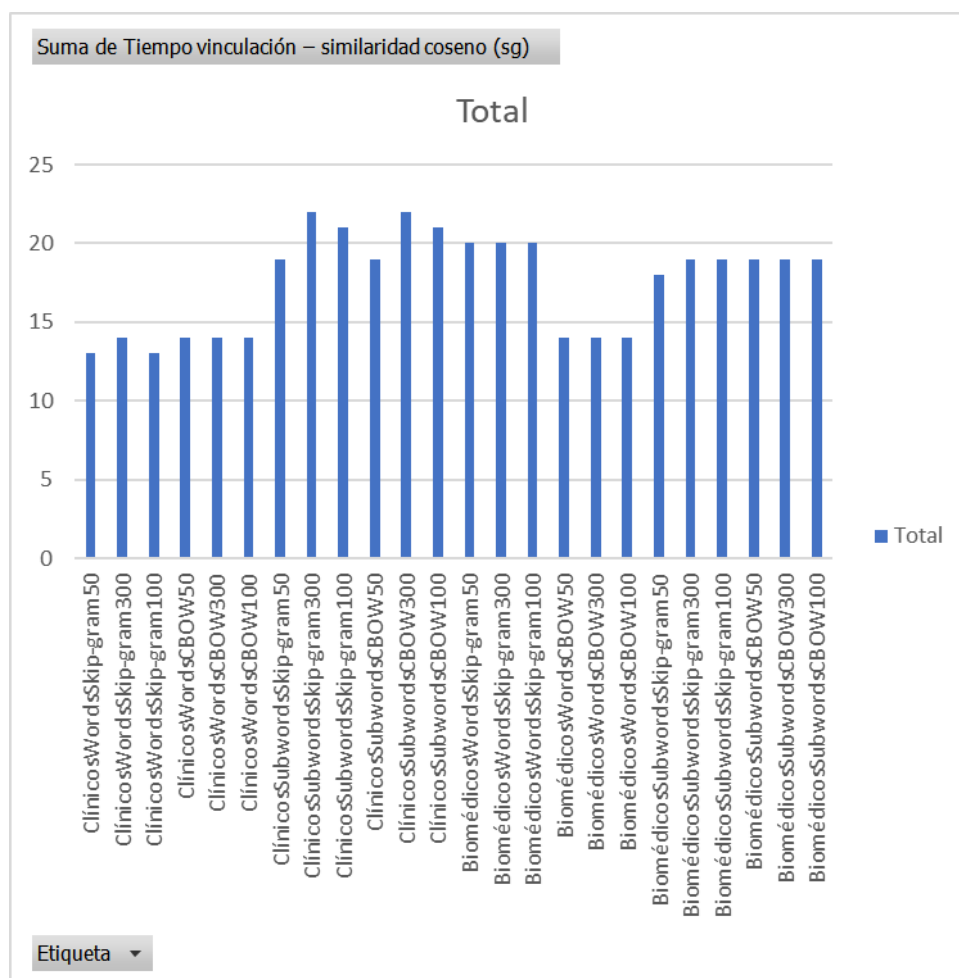


Ilustración 4. Comparación de tiempos de vinculación del informe tipo mediante similitud coseno en cada uno de los modelos. Elaboración propia.

Se observa que el tiempo de vinculación del informe tipo utilizando la similitud coseno como medida es muy similar en todos los modelos, aunque los modelos los modelos que trabajan a nivel de palabras parecen ser algo más rápidos en general.

Por otro lado, en cuanto a los umbrales, se han determinado empíricamente para tratar de encontrar unos umbrales que obtengan resultados aceptables, y comparables entre sí. Así, se observa que el modelo mejora en lo que respecta a la detección de términos vinculados, conforme se aumenta el número de dimensiones de los vectores densos. El modo de entrenamiento no afecta en la mejora de la vinculación.

- **TFM\_Benchmarking\_Transformers v1.ipynb:**

Cuaderno desarrollado para comprobar los resultados del modelo basado en *Transformers*, “BETO-Galén”, entrenado con textos clínicos en español, en este caso utilizado para la obtención de vectores densos.

El preprocesado del texto se realiza con la librería *spacy*, con el modelo más extenso disponible para el preprocesado y división en frases de los textos (<https://spacy.io/models/es>).

Se cargan los catálogos de terminologías clínicas indicadas (problemas de salud en el ámbito hospitalario), con el modelo disponible, guardando en fichero los vectores densos de todos los términos clínicos para agilizar la ejecución, tomando tiempos de cada paso del proceso.

Se carga el modelo *BETO-Galén*, con la librería *keras\_bert* (<https://pypi.org/project/keras-bert/>), para la generación de vectores densos de 768 dimensiones.

Después se ejecuta, con este modelo, el algoritmo de obtención de términos clínicos relacionados, para un informe radiológico tipo, con el catálogo precargado previamente. De estas operaciones se toman tiempos y se comparan los resultados con los esperados para dicho informe médico.

Se destaca que para la similitud entre palabras o grupo de palabras se utilizan dos tipos de medidas, la distancia euclidiana y la similitud coseno.

Después de analizar el cuaderno, se constata que la carga del modelo es de 60 segundos, que la generación de vectores para el catálogo de terminologías clínicas es de 40 minutos (medio minuto por término aproximadamente), el tiempo de vinculación con el informe tipo es de 16 segundos con la distancia euclidiana o de 19 segundos con similitud coseno.

El resultado obtiene 24 términos clínicos para el informe médico modelo, con muchos términos que no guardan relación, por lo que no se considera para su uso dentro del motor de inferencia.

## Apéndice II

### Desarrollo del motor de inferencia

Como se ha mencionado previamente la infraestructura tecnológica del motor, es un entorno Python 3.8, con múltiples librerías para la funcionalidad perseguida, con base de datos *MySQL Community Server 8.0* para la persistencia de datos.

La arquitectura del software está basada en capas (multicapa), separando y agrupando las funciones en cada una de ellas:

- 1) Capa de entrada (API REST).
- 2) Capa de gestores (*managers*).
- 3) Capa de acceso a datos.
- 4) Capa general de utilidades.

- **Capa de entrada (API REST)**

La capa de entrada contiene un módulo con funciones que gestionan las operaciones de solicitud y entrega de información a través de un *API REST*, encargándose únicamente de la validación de los datos de entrada, y de la sesión de los usuarios. Esta capa utiliza la siguiente capa, de gestores, para ejecutar acciones u obtener la información solicitada.

El módulo se denomina *ClinicalNELEngineREST.py*, y las operaciones implementadas actualmente se indican a continuación.

operación	URL	parámetros	acción	retorno
GET	/application	user	Obtener la versión del motor.	Versión del motor.
GET	/users	user, password	Validar al usuario (contra directorio activo).	Token de acceso.
PUT	/terminosclinicos	user, token, idtermino, termino	Alta de nuevo término clínico.	Término clínico creado.
POST	/terminosclinicos	user, token	Inicialización y carga del catálogo de términos clínicos.	-
GET	/terminosclinicos	user, token, id (opcional), termino (opcional)	Consultar los términos clínicos disponibles en el motor.	Términos clínicos.
PUT	/reports	user, token, reportid, informe	Creación de nuevo informe clínico.	Términos clínicos rela-

		(body)		cionados.
<b>GET</b>	/reports	user, token, idtermino (opcional), report_id (opcional), owner (opciones)	Consulta de los informes disponibles o filtrados por un término clínico o un identificador.	Informes clínicos.

*Tabla 10. Operaciones implementadas en el motor de inferencia desarrollado. Elaboración propia.*

Las operaciones mencionadas serán invocadas por el prototipo de interfaz de usuario, de forma transparente, aunque también pueden ser invocadas mediante una herramienta externa como el *Postman* (<https://www.postman.com/>), para la realización de pruebas y depuración.

Se debe tener en cuenta en la llamada, incluir la *URL* donde esté arrancado el motor de inferencia (típicamente <http://127.0.0.7:5003>), utilizar el método adecuado (*GET*, *PUT*, *POST*), incluir todos los parámetros indicados en la cabecera (*header*), salvo el cuerpo del informe que irá en el cuerpo de la invocación (*body*). El resultado se devuelve en formato *json* del tipo: `{'data': '<datos_resultado>', 'result': 'success'}` o bien `{'result': 'error', 'message': '<mensaje de error>'}`. Es particularmente importante indicar que es necesario validar al usuario para obtener el *token*, que debe ser utilizando posteriormente en el resto de llamadas al motor.

- **Capa de gestores (managers)**

Esta capa contiene varios módulos con la lógica de negocio del motor de inferencia, es la que orquesta las instrucciones a llevar a cabo a alto nivel del motor. Esta capa utiliza las siguientes capas para el acceso a la base de datos, utilidades generales o el procesamiento del lenguaje natural.

Todos los módulos recogen el nombre de usuario, y la dirección IP desde la que se hace la invocación (al *API REST*), para llevar la traza/registro de operaciones.

### **Módulo *TerminosClinicosManager.py***

Este módulo contiene las funciones con la lógica de negocio a aplicar sobre las terminologías clínicas implementadas para dar servicio al *API REST* del motor de inferencia.

función	parámetros	acción	retorno
se-	User, remot_addr,	Obtener un tér-	Término

<b>lect_terminoclinico</b>	id	mino clínico en base a su código.	clínico.
<b>se-lect_terminoclinico_by_id</b>	User, remot_addr, id	<b>Obtener un término clínico en base a su identificador interno.</b>	<b>Término clínico.</b>
se-lect_terminosclnicos	User, remot_addr, id (opcional), termino (opcional), num_palabras (opcional)	Obtener todos los términos clínicos filtrados por los criterios.	Lista de términos clínicos.
<b>create_terminoclinico</b>	User, remot_addr, idtermino, termino	<b>Crear un nuevo término clínico.</b>	<b>Identificador del término clínico creado.</b>
update_terminoclinico	User, remot_addr, idtermino, termino	Modificar un término clínico existente.	Identificador del término clínico modificado.
<b>delete_terminoclinico</b>	User, remot_addr, id	<b>Borrar un término clínico existente.</b>	<b>Identificador del término clínico borrado.</b>
Load_terminosclnicos	User, remot_addr	Recargar los términos clínicos en base a un fichero de texto de partida, calculando el <i>embedding</i> para cada uno de ellos.	-

Tabla 11. Funciones implementadas en el módulo de gestión de términos clínicos. Elaboración propia.

### Módulo *ReportsManager.py*

Este módulo contiene las funciones con la lógica de negocio a aplicar sobre los informes médicos implementadas para dar servicio al *API REST* del motor de inferencia.

función	parámetros	acción	retorno
<b>select_report</b>	User, remot_addr,	Obtener un infor-	Informe

	id	me en base a su identificador.	médico.
<b>select_reports</b>	<b>User, remot_addr</b>	<b>Obtener todos los informes clínicos.</b>	<b>Lista de informes médicos.</b>
create_report	User, remot_addr, reportid, re- portbody, model (opcional)	Crear un nuevo informe médico, y asignación de eti- quetas de la ter- minología.	Términos clínicos etiqueta- dos.
<b>upda- te_terminoclinico</b>	<b>User, remot_addr, idtermino, termino</b>	<b>Modificar un tér- mino clínico exis- tente.</b>	<b>Identifica- dor del término clínico mo- dificado.</b>
dele- te_terminoclinico	User, remot_addr, id	Borrar un término clínico existente.	Identifica- dor del término clínico bo- rrado.
<b>get_reports</b>	<b>User, remot_addr, idinforme, propie- tario, termino</b>	<b>Obtener los in- formes médicos filtrados por el idinforme, el pro- pietario o el tér- mino suministra- do.</b>	<b>Lista de informes médicos.</b>
Get_reports_and_la bels	User, remot_addr, idinforme, propie- tario, termino	Obtener los infor- mes médicos fil- trados por el idin- forme, el propieta- rio o el término suministrado.	Lista de informes médicos junto con sus etique- tas.
<b>get_termino_report s</b>	<b>User, remot_addr, reportid</b>	<b>Obtener los tér- minos clínicos relacionados con un informe médi- co especificado.</b>	<b>Lista de términos clínicos relaciona- dos.</b>
Get_terminosclnicos _FastText	User, informe (cuerpo)	Lleva a cabo la obtención de eti- quetas con el mo- delo basado en FastText	Lista de términos clínicos relaciona- dos.

Tabla 12. Funciones implementadas en el módulo de gestión de informes. Elaboración propia.



Este módulo contiene la lógica clave para conseguir uno de los objetivos del TFM, que es experimentar con un modelo para etiquetar los informes con términos relacionados en base a una terminología clínica.

### Módulo *Users.py*

Este módulo contiene las funciones con la lógica de negocio a aplicar sobre los usuarios para dar servicio al *API REST* del motor de inferencia.

función	parámetros	acción	retorno
<b>validateUser</b>	User, password remot_addr	Validar al usuario en base a su nombre de usuario y contraseña.	Verdadero o falso.

Tabla 13. Funciones implementadas en el módulo de usuarios. Elaboración propia.

- **Capa de acceso a datos**

Esta capa contiene módulos con la lógica necesaria para gestionar los datos contra la base de datos, se encarga de las operaciones CRUDS, de creación, recuperación, actualización, borrado y consulta de los datos de la base de datos.

Se utilizan las librerías de conexión contra base de datos *mysql.connector 2.2.9* (<https://pypi.org/project/mysql-connector/>).

### Módulo *TerminosClinicosDAO.py*

Este módulo contiene las funciones de la capa de persistencia en base de datos a aplicar sobre las terminologías clínicas, implementadas para dar servicio al *API REST* del motor de inferencia.

función	parámetros	acción	retorno
<b>select_terminoclinico</b>	idtermino	Obtener de base de datos un término clínico en base a su identificador.	Término clínico.

<b>se- lect_terminosclini- cos</b>		<b>Obtener de base de datos todos los términos clínicos.</b>	<b>Lista de términos clínicos.</b>
crea- te_terminoclinico	idtermino, termino, embedding	Crear un nuevo término clínico en la base de datos.	Identifica- dor del término clínico creado.
<b>upda- te_terminoclinico</b>	<b>idtermino, termi- no, embedding</b>	<b>Modificar en la base de datos un término clínico existente.</b>	<b>Verdade- ro/falso</b>
dele- te_terminoclinico	id	Borrar un término clínico existente en la base de da- tos.	
<b>dele- te_terminosclnicos</b>		<b>Borrar todos los términos clínico existentes de la base de datos.</b>	

Tabla 14. Funciones implementadas en el módulo de acceso a datos de términos clínicos. Elaboración propia.

### Módulo *ReportsDAO.py*

Este módulo contiene las funciones de la capa de persistencia en base de datos a aplicar sobre los informes médicos, implementadas para dar servicio al *API REST* del motor de inferencia.

<b>función</b>	<b>parámetros</b>	<b>acción</b>	<b>retorno</b>
<b>select_report</b>	id	Obtener de la base de datos un informe en base a su identificador y propietario.	Informe médico.
<b>select_reports</b>		<b>Obtener todos los informes clínicos de la base de datos.</b>	<b>Lista de informes médicos.</b>
create_report	idreport, re- portbody	Crear un nuevo informe médico en la base de datos.	Identifica- dor del in- forme mé- dico creado.
<b>update_report</b>	<b>idreport, re-</b>	<b>Modificar un in-</b>	<b>-</b>

	<b>portbody</b>	<b>forme médico existente en la base de datos.</b>	
delete_report	reportid	Borrar un informe médico existente.	-
<b>get_reports</b>	<b>idtermino</b>	<b>Obtener de la base de datos los informes médicos que contienen el término clínico indicado.</b>	<b>Lista de informes médicos.</b>
select_termino_reports	Report_id	Obtener de la base de datos los términos clínicos relacionados con un informe médico especificado.	Lista de términos clínicos relacionados.
<b>Create_termino_report</b>	<b>Report_id, idtermino, similarity, model</b>	<b>Asociar un término clínico a un informe médico en la base de datos.</b>	<b>Identificador del término clínico asociado</b>
Delete_termino_reports	Report_id	Eliminar de la base de datos los términos clínicos asociados a un informe concreto.	-

Tabla 15. Funciones implementadas en el módulo de acceso a datos de informes.  
 Elaboración propia.

- Capa general de utilidades**

Esta capa contiene módulos con la lógica necesaria para utilidades generales del motor (configuración, registro de valores constantes, criptografía, excepciones, conexión contra directorio activo, traza de la aplicación, utilidades, utilidades PLN).

### **Módulo *configuration.py***

Este módulo contiene los parámetros de configuración del motor, como son la versión y fecha de revisión, la ruta al servidor de directorio activo (para validación), el usuario/contraseña y acceso a la base de datos, la configuración, dirección *IP* y puerto del servidor *API REST*.

### **Módulo *constants.py***

Este módulo contiene los valores constantes a utilizar en el motor, como son los literales a utilizar en las trazas de la aplicación y los posibles modelos a utilizar.

### **Módulo *crypto.py***

Este módulo contiene las operaciones criptológicas a utilizar en el motor, como son la generación y verificación de los *tokens* de autenticación de los usuarios.

Se utilizan la librería de criptografía *itsdangerous* 2.1.2 (<https://pypi.org/project/itsdangerous/>).

### **Módulo *ldap.py***

Este módulo contiene las operaciones de conexión contra directorio activo a utilizar en el motor, como son la autenticación de los usuarios.

Se utiliza la librería de conexión contra el directorio activo *ldap3* 2.9.1 (<https://pypi.org/project/ldap3/>).

### **Módulo *logs.py***

Este módulo contiene las operaciones de registro de traza en disco a utilizar en el motor, como son el registro de acciones general y el registro de acciones relacionadas con acceso a datos personales (para mantener la trazabilidad de los accesos).

### **Módulo *NLPutils.py***

Este módulo contiene las funciones de procesamiento del lenguaje natural a utilizar en el motor, como son el preprocesado de textos, la obtención de *embeddings*, el cálculo de similitud entre vectores densos, la obtención de grupos de n-gramas de un texto, la división en frases/sentencias y la carga de los modelos asociados.

Se utilizan las librerías *gensim* 4.3 (<https://pypi.org/project/gensim/>) para la gestión de modelos Fasttext, *spacy* 3.5 (<https://pypi.org/project/spacy/>) para la gestión del preprocesado del texto y el modelo del lenguaje, *pandas* 1.5 (<https://pypi.org/project/pandas/>), *numpy* 1.24 (<https://pypi.org/project/numpy/>) para trabajar con estructuras de datos avanzadas.

### **Módulo *utils.py***

Este módulo contiene las funciones generales a utilizar en el motor, como son la serialización/deserialización de listas de objetos, que se utilizan para poder almacenar en base de datos los *embeddings* de los términos clínicos.

- **Módulo de PLN**

Aunque este módulo está dentro de la capa general de utilidades (**NLPutils.py**), se presenta en este punto para dar mayor detalle sobre la misma.

Las funciones implementadas en este módulo son las siguientes:

### **loadmodel\_NLP**

Carga del modelo utilizado por la librería *spacy*, en este caso el modelo “*es\_core\_news\_lg*” (<https://spacy.io/models/es>), que es el modelo extenso en español, entrenado a base de noticias generales en diferentes medios, que incluye un etiquetador, un vectorizador de *tokens*, una herramienta de morfología, un analizador del texto, la posibilidad de establecer reglas sobre atributos, un lematizador, un reconocedor de entidades nombradas, y una lista de vectores de palabras con 500.000 vectores únicos (con 300 dimensiones).

### **loadmodel\_FastText**

Carga del modelo, en este caso el modelo “*Biomedical-Word-Embedding-for-Spanish*” (<https://github.com/PlanTL-GOB-ES/Biomedical-Word-Embeddings-for-Spanish>), en este caso un modelo que genera *FastText embeddings* de 300 dimensiones, entrenado mediante *CBOW*, en base a casos clínicos.

Este recurso contiene varios modelos en español, entrenados mediante *CBOW* y *Skip-gram*, sensibles al caso (mayúsculas/minúsculas) y también independientes de él, un grupo de ellos entrenados a base de recursos médicos, literatura científica, casos clínicos y otros datos, y otro grupo de ellos entrenados en base a casos clínicos, principalmente de COVID-19 y de ictus. Los modelos están entrenados para obtener una lista de vectores de palabras o *FastText embeddings*, a nivel de palabras completas, **Word embeddings**, y a nivel de subpalabras **Byte-Pair Encoding (BPE) subword embeddings** con un vocabulario de 8.000/10.000 *tokens* (de 50, 100 y 300 dimensiones).

### **preprocesaDoc**

Función que transforma el texto utilizando para ello la librería *Spacy*, el modelo cargado previamente de esta librería, con el fin de realizar un proceso que lo prepare para las operaciones posteriores.

Las operaciones de preprocesado realizadas son las siguientes:

- Conversión del texto a minúsculas (independiente del caso).
- Tokenizado del texto en palabras.
- Eliminación de signos de puntuación del texto.
- Eliminación de palabras irrelevantes (*stopwords*) del texto.
- Lematización de palabras.
- Eliminación de tilde o diéresis (sustituyendo por caracteres sin estos signos).
- Eliminación de espacios previos/posteriores.

## getEmbedding

Función que obtiene el vector denso (*embedding*) de un palabra o grupo de palabras utilizando el modelo cargado de obtención de *embeddings* mencionado.

Para ello, se obtiene el vector denso de cada una de las palabras por separado y después los vectores se promedian para obtener un único vector denso que represente el grupo de palabras.

## calculateCosineSimilarity

Función que calcula la distancia o similitud coseno entre dos vectores densos, mediante una operación matemática específica.

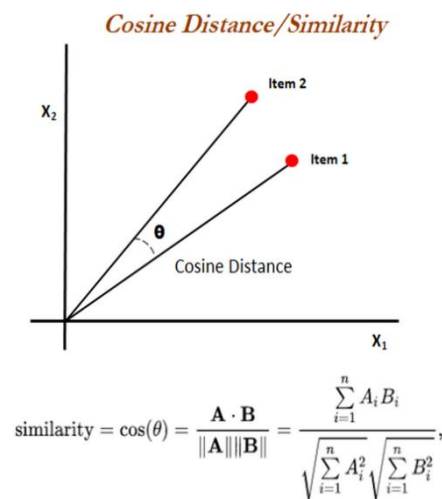


Ilustración 5. Representación gráfica y fórmula matemática para obtener la distancia o similitud coseno.

Fuente: <https://www.tyrrell4innovation.ca>

Esta similitud coseno es la que representa cómo de similares son dos palabras o grupos de palabras, en base a la orientación de sus vectores densos (*embeddings*). Si el resultado de esta operación es “1”, esto significa que los vectores densos tienen la misma dirección, totalmente similares, mientras que, si el resultado de esta operación es “0”, las direcciones de los vectores densos forman un ángulo recto, totalmente distintos (*Cosine Similarity, Cosine Distance Explained | Math, Statistics for Data Science, Machine Learning - YouTube*, n.d.).

## calculateEuclideanDistance

Función que calcula la distancia euclídea entre dos vectores densos, mediante una operación matemática específica.

$$\|\vec{AB}\| = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2} \text{ siendo } \vec{OA} = (a_1, a_2, \dots, a_n) \text{ y } \vec{OB} = (b_1, b_2, \dots, b_n).$$

$$\|\mathbf{v}\| = \sqrt{v_1^2 + v_2^2 + \dots + v_n^2} = \sqrt{\sum_{i=1}^n v_i^2}$$

*Ilustración 6. Fórmula matemática para obtener la distancia euclidiana o norma euclídea. Fuente:*  
<https://es.wikipedia.org>

La norma de un vector es la distancia euclídea (en línea recta) entre dos puntos A y B que delimitan dicho vector.

Esta distancia euclídea es la que representa cómo de similares son dos palabras o grupos de palabras, en base a la distancia euclídea de sus vectores densos (*embeddings*). Si el resultado de esta operación es “0”, esto significa que los valores de los vectores densos son iguales, totalmente similares, mientras que, si el resultado de esta operación es mayor, las distancias entre los valores de los vectores densos aumentan, distintos/similares en mayor o menor medida.

### **NLP\_get\_ngrams**

Función que convierte un texto en una lista de grupos de palabras consecutivas del texto de diferentes tamaños (cada grupo desde una hasta un número determinado de palabras).

### **NLP\_split\_sentences**

Función que convierte un texto en una lista de frases contenidas en dicho texto, utilizando el modelo cargado para la librería *Spacy*.

El código fuente está incluido en el repositorio de *github* correspondiente:  
<https://github.com/fmedinafernandez/14MBID-TFM-NEL>

## Apéndice III

### Base de Datos

A continuación, se presenta el diagrama de entidad relación y los datos almacenados en la base de datos del prototipo. La base de datos tiene una estructura muy simple, cuya función principal es la de apoyar las funcionalidades del prototipo, esto es, el almacenamiento y recuperación de los datos del catálogo de los términos clínicos y de los informes médicos junto con sus términos clínicos asociados reconocidos por el motor.

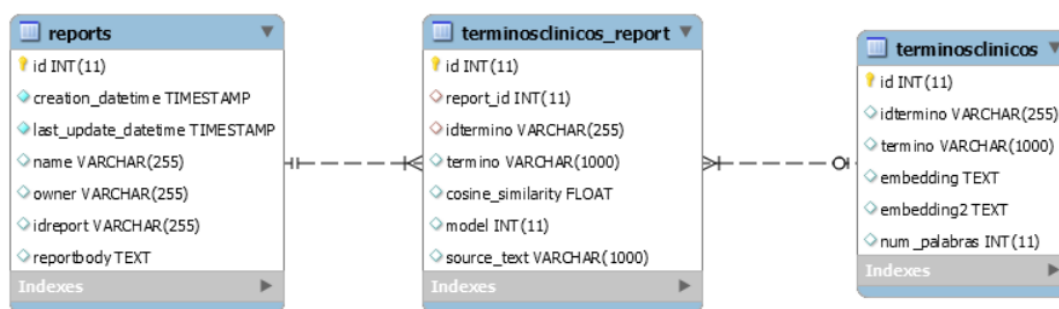


Ilustración 7. Diagrama físico de la base de datos. Elaboración propia mediante MySQL Workbench.

La base de datos del prototipo almacena la siguiente información:

- **Términos clínicos.** Catálogo de términos clínicos que se utilizan para la vinculación con los informes médicos, junto con los vectores densos obtenidos en base al modelo utilizado.
- **Informes.** Repositorio de informes médicos que se han procesado mediante el prototipo, incluyendo el cuerpo del informe como información principal.
- **Términos clínicos asociados/enlazados.** Repositorio de términos clínicos asociados a los informes médicos procesados, incluyendo la similitud que se ha determinado en base al modelo y el algoritmo propuesto.

El modelo físico implementado en MySQL Community Server 8.0 (<https://dev.mysql.com/downloads/mysql/>) contiene las siguientes tablas y campos que se indican a continuación:



tabla	campo	tipo	comentario
<b>terminosclini- cos</b>	Id (clave primaria)	int	Identificador único del término clínico.
	<b>idtermino</b>	<b>Varchar(255)</b>	<b>Identificador o código del término.</b>
	termino	Var- char(1000)	Término clínico (lite- ral).
	<b>Num_palabras</b>	<b>int</b>	<b>Número de palabras del término clínico (preprocesado)</b>
	embedding	text	Vector denso ( <i>embed- ding</i> ) que representa al término clínico (se- rializado en formato <i>json</i> ).
	<b>embedding2</b>	<b>text</b>	<b>Vector denso alterna- tivo (<i>embedding</i>) que representa al término clínico (serializado en formato <i>json</i>).</b>
reports	Id (clave primaria)	int	Identificador único del informe.
	<b>Creation_datetime</b>	<b>Timestamp</b>	<b>Marca de tiempo de creación del informe.</b>
	Last_update_dateti me	timestamp	Marca de tiempo de última modificación del informe.
	<b>name</b>	<b>Varchar(255)</b>	<b>Nombre del informe.</b>
	owner	Varchar(255)	Nombre del propietario del informe.
	<b>idreport</b>	<b>Varchar(255)</b>	<b>Identificador o código del informe.</b>
	reportbody	Text	Informe médico (texto o cuerpo del informe).
<b>Terminosclini- cos_report</b>	<b>Id</b>	<b>int</b>	<b>Identificador único del término clínico asociado.</b>
	Report_id	Int	Identificador único del informe relacionado.
	<b>Source_text</b>	<b>Var- char(1000)</b>	<b>Parte del texto origen preprocesado similar al término vinculado</b>
	Idtermino	Varchar(255)	Identificador o código del término.
	<b>Termino</b>	<b>Var- char(1000)</b>	<b>Término clínico (lite- ral).</b>

	Cosine_similarity	Float	Medida de similitud coseno.
	<b>model</b>	<b>Int</b>	<b>Identificador del modelo utilizado para generar los vectores densos.</b>

*Tabla 16. Tablas, campos y tipos de datos de la base de datos. Elaboración propia.*

## Apéndice IV

### Desarrollo del interfaz de usuario prototipo de vinculación de informes médicos

Como se ha mencionado previamente el lenguaje de maquetado es HTML, se utilizan hojas de estilo CSS para gestionar la apariencia y código *Javascript* para que el interfaz sea interactivo y se integre con el motor de integración. El servidor de aplicaciones se ha implementado utilizando una librería desarrollada en *Python*.

- **Interfaz de usuario y menú de navegación**

A continuación, se muestra la pantalla inicial de la aplicación donde se dispone de un menú en la parte superior con acceso a las diferentes funcionalidades de la aplicación.



*Ilustración 8. Página de inicio del prototipo. Elaboración propia.*

Desde esta pantalla inicial se puede acceder a los siguientes menús:

1. **Menú “Inicio”**. Esta opción de menú muestra la pantalla inicial de bienvenida a la aplicación.
2. **Menú “Informes”**. Esta opción de menú muestra las opciones disponibles con respecto a los informes médicos. En este caso es posible la creación y etiquetado automático de un nuevo informe y la consulta de informes creados previamente en base a las etiquetas asignadas.
3. **Menú “Terminología”**. Esta opción de menú muestra las opciones disponibles con respecto a la terminología clínica. En este caso es posible la creación de un nuevo término clínico y la consulta de los términos clínicos cargados previamente (enfermedades más habituales en hospitalización de la terminología SNOMED-CT en español).
4. **Menú “Acerca de”**. Esta opción de menú muestra la pantalla con información adicional de la aplicación.

El acceso a las diferentes pantallas y funcionalidades de la aplicación se hace mediante un “clic” con el botón derecho del ratón en la opción correspondiente del menú. Algunos menús disponen de un menú secundario para el acceso a funcionalidades relacionadas.

- **Interfaz de inicio de la aplicación**

A continuación, se muestra la pantalla de la aplicación donde se presenta la versión de la aplicación.



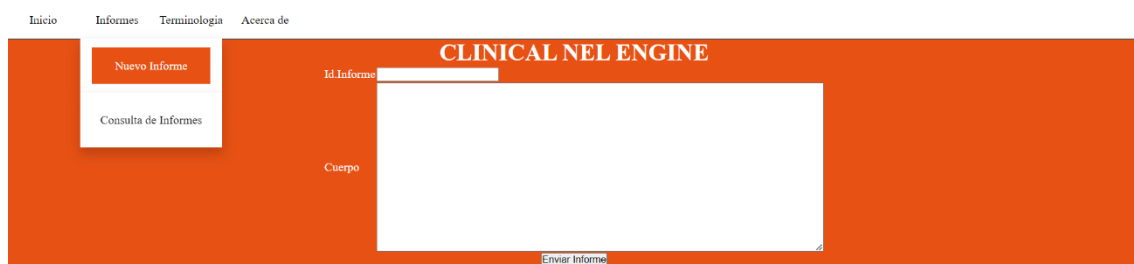
- **Interfaz de gestión de informes médicos**

A continuación, se muestra la pantalla de la aplicación donde se dispone de un menú en la parte superior con acceso a las diferentes funcionalidades relativas a los informes médicos de la aplicación.



*Ilustración 9. Menú de informes del prototipo. Elaboración propia.*

Para la creación y etiquetado de nuevos informes se dispone de un menú secundario “Nuevo Informe” que muestra una pantalla con una caja de texto para incluir el cuerpo del informe a crear y un botón para llevar a cabo la creación.



*Ilustración 10. Interfaz de creación y etiquetado de informes del prototipo. Elaboración propia.*

Una vez se pulsa el botón el informe es creado en el servidor y se detona la ejecución del motor PLN para obtener y vincular las etiquetas de la terminología clínica que se determina. La aplicación devuelve interactivamente las etiquetas vinculadas al informe en un tiempo de respuesta razonable que depende del tamaño del informe.

Para la consulta de informes se dispone de un menú secundario “Consulta de Informes” que muestra una pantalla con varias cajas de texto para introducir los filtros de búsqueda del informe y un botón para llevar a cabo la consulta de informes.

A continuación, se muestra un informe tipo junto con la información obtenida:

### CLINICAL NEL ENGINE

**Id.Informe**

**Cuerpo**

ID24675

No hay evidencia de obstrucción del intestino delgado o grueso.

No hay evidencia de ascitis abdominal o linfadenopatía.

No hay evidencia de masa vesical intrínseca o extrínseca .

No hay ascitis pélvica ni linfadenopatía .

El útero y los ovarios no presentan ninguna anomalía.

Las imágenes de las bases pulmonares no muestran evidencia de masa pleural o parenquimatosa .

No hay derrames pleurales.

Hay cicatrices en el lóbulo medio derecho y en la lingula, así como en ambas bases pulmonares.

Las estructuras óseas están libres de lesiones líticas o blásticas.

Se observan cambios degenerativos multinivel en la columna toracolumbar.

Se observan calcificaciones dispersas en la aorta y en sus principales ramas, compatibles con la aterosclerosis.

[Enviar Informe](#)

Fuente preprocesada	Término	Distancia/similitud
apendicitis	apendicitis	100%
ascitis	ascitis	100%
lingular base	bronquiectasias en ambas bases	92.4848%
cambio degenerativo multinivel	cambios degenerativos en columna	90.24560000000001%
cicatriz	cicatriz	100%
calculo renal	cálculo renal	100%
derrame pleural	derrame pleural	100%
dilatacion conducto biliar	dilatación del conducto de Wirsung	90.6955%

*Ilustración 11. Resultado del etiquetado de un nuevo informe. Elaboración propia.*

Inicio
Informes
Terminología
Acercas de

Nuevo Informe

Consulta de Informes

### CLINICAL NEL ENGINE

Id informe
Propietario
Informes
Término clínico
Buscar Informes

*Ilustración 12. Interfaz de consulta de informes del prototipo. Elaboración propia.*

**Filtro “Id informe”:** Este filtro busca los informes cuyo identificador de informe contiene el texto introducido.

**Filtro “Propietario”:** Este filtro busca los informes cuyo propietario coincide con el texto introducido.

**Filtro “Término clínico”:** Este filtro busca los informes cuyas etiquetas contienen el texto introducido.

Una vez se pulsa el botón el servidor consulta los informes médicos disponibles que cumplen dichos filtros de búsqueda. La aplicación devuelve interactivamente los informes que cumplen los criterios.

A continuación, se muestra el resultado de una búsqueda de informes junto con la información obtenida de etiquetas:

Identificador	Usuario	Cuerpo	Etiquetas																																							
id0001	usuario	Hay evidencia de hipoatenuación hepática difusa compatible con infiltración grasa. No hay dilatación de los conductos biliares intra o extrahepáticos. El paciente se encuentra en estado post colecistectomía. El bazo es normal. El páncreas es de contorno y características de atenuación normales. No hay evidencia de masa suprarrenal. Hay una hernia supraumbilical de tamaño moderado que contiene grasa. Los riñones son normales en tamaño, forma y configuración. No se identifican cálculos renales ni ureterales. No hay hidroureter ni hidronefrosis. No hay evidencia de apendicitis. Hay varias asas de intestino delgado llenas de líquido, compatibles con una enteritis leve. No hay engrosamiento de la pared intestinal. No hay evidencia de obstrucción del intestino delgado o grueso. No hay evidencia de ascitis abdominal o linfadenopatía. No hay evidencia de masa vesical intrínseca o extrínseca. No hay ascitis pélvica ni linfadenopatía. El útero y los ovarios no	<table><tr><th>Fuente preprocesada</th><th>Término</th><th>Distancia/similitud</th></tr><tr><td>apendicitis</td><td>apendicitis</td><td>100%</td></tr><tr><td>ascitis</td><td>ascitis</td><td>100%</td></tr><tr><td>lingular base</td><td>bronquiectasias en ambas bases</td><td>92.4848%</td></tr><tr><td>cambio degenerativo multinivel</td><td>cambios degenerativos en columna</td><td>90.24560000000001%</td></tr><tr><td>cicatriz</td><td>cicatriz</td><td>100%</td></tr><tr><td>calculo renal</td><td>cálculo renal</td><td>100%</td></tr><tr><td>derrame pleural</td><td>derrame pleural</td><td>100%</td></tr><tr><td>dilatacion conducto biliar</td><td>dilatación del conducto de Wirsung</td><td>90.6955%</td></tr><tr><td>dilatacion conducto biliar intro</td><td>dilatación, incluso quística, de los ductos biliares</td><td>90.9952%</td></tr><tr><td>asa intestino delgado lleno</td><td>edemas en asas de intestino delgado</td><td>90.2751%</td></tr><tr><td>enteritis leve</td><td>escleritis leve</td><td>93.3063%</td></tr><tr><td>degenerativo multinivel</td><td>espondilolistesis degenerativa</td><td>90.984%</td></tr></table>	Fuente preprocesada	Término	Distancia/similitud	apendicitis	apendicitis	100%	ascitis	ascitis	100%	lingular base	bronquiectasias en ambas bases	92.4848%	cambio degenerativo multinivel	cambios degenerativos en columna	90.24560000000001%	cicatriz	cicatriz	100%	calculo renal	cálculo renal	100%	derrame pleural	derrame pleural	100%	dilatacion conducto biliar	dilatación del conducto de Wirsung	90.6955%	dilatacion conducto biliar intro	dilatación, incluso quística, de los ductos biliares	90.9952%	asa intestino delgado lleno	edemas en asas de intestino delgado	90.2751%	enteritis leve	escleritis leve	93.3063%	degenerativo multinivel	espondilolistesis degenerativa	90.984%
			Fuente preprocesada	Término	Distancia/similitud																																					
			apendicitis	apendicitis	100%																																					
			ascitis	ascitis	100%																																					
			lingular base	bronquiectasias en ambas bases	92.4848%																																					
			cambio degenerativo multinivel	cambios degenerativos en columna	90.24560000000001%																																					
			cicatriz	cicatriz	100%																																					
			calculo renal	cálculo renal	100%																																					
			derrame pleural	derrame pleural	100%																																					
			dilatacion conducto biliar	dilatación del conducto de Wirsung	90.6955%																																					
			dilatacion conducto biliar intro	dilatación, incluso quística, de los ductos biliares	90.9952%																																					
			asa intestino delgado lleno	edemas en asas de intestino delgado	90.2751%																																					
			enteritis leve	escleritis leve	93.3063%																																					
			degenerativo multinivel	espondilolistesis degenerativa	90.984%																																					

Ilustración 13. Resultado de la consulta de informes. Elaboración propia.

- **Interfaz de gestión de términos clínicos**

A continuación, se muestra la pantalla de la aplicación donde se dispone de un menú en la parte superior con acceso a las diferentes funcionalidades relativas a los términos clínicos de la aplicación.



Ilustración 14. Menú de terminología clínica del prototipo. Elaboración propia.

Para la creación de nuevos términos clínicos se dispone de un menú secundario “Nuevo Término” que muestra una pantalla con dos cajas de texto para incluir el

código y la descripción del término clínico a crear y un botón para llevar a cabo la creación.

*Ilustración 15. Interfaz de creación de términos clínicos del prototipo. Elaboración propia.*

Una vez se pulsa el botón el término clínico es creado en el servidor y se detona la ejecución del motor PLN para obtener el vector denso asociado (*embedding*).

Para la consulta de términos clínicos se dispone de un menú secundario “Consulta de Términos” que muestra una pantalla con varias cajas de texto para introducir los filtros de búsqueda de los términos y un botón para llevar a cabo la consulta de estos.

*Ilustración 16. Interfaz de consulta de términos clínicos del prototipo. Elaboración propia.*

**Filtro “Código término”:** Este filtro busca los términos cuyo código contiene el texto introducido.

**Filtro “Término”:** Este filtro busca los términos cuya descripción coincide con el texto introducido.

Una vez se pulsa el botón el servidor consulta los términos clínicos disponibles que cumplen dichos filtros de búsqueda. La aplicación devuelve interactivamente los términos clínicos que cumplen los criterios.

CLINICAL NEL ENGINE		
<div> <div>Código término</div> <div> <div>Terminos Clínicos</div> <div> <div>Termino</div> <div> <div>Buscar Términos Clínicos</div> </div> </div> </div> </div>		
Identificador	Código	Descripción
11037	b6e59780-57c9-4f02-ccb3-363fbc74bed	abdomen distendido
11038	921ef527-d0ff-40a3-8c10-0e8189f2206f	aborto espontáneo
11039	29860941-a1df-4ae6-8ec2-a45630077dff	aborto espontáneo completo
11040	6b9fcb5e-5e34-4eba-8436-d2ec66dd5bb7	aborto espontáneo incompleto
11041	b5175825-ed45-4cf3-8a65-175586499e8	aborto inadvertido
11042	5229972a-2a70-463b-8db7-db47a2d5d081	aborto retenido
11043	b8e9fc56-2f30-47dd-ae05-1cd98efa5842	abrasión
11044	e20c4d77-df7a-4494-9a13-b7b65c413b34	abrasión corneal
11045	c4d07f0a-9d38-4a36-9571-e189437db543	abrasión de extremidad inferior
11046	ac952f57-bc47-4709-8dd1-845944ff93a	absceso
11047	d421ff0a-c13d-4de8-b7f6-1243d3068551	absceso abdominal
11048	51cc905c-98b0-46ee-b112-5558993ae974	absceso de amígdala
11049	f86b43f1-6707-4d38-9f2a-cdce4d2a0e9b	absceso de hígado

Ilustración 17. Interfaz de visualización de términos clínicos del prototipo. Elaboración propia.

- **Interfaz de “acerca de” de la aplicación**

A continuación, se muestra la pantalla de la aplicación donde se dispone de un menú en la parte superior con acceso a la funcionalidad de “Acerca de”.



Ilustración 18. Menú de “acerca de” del prototipo. Elaboración propia.

La estructura del interfaz de usuario contiene los siguientes módulos.

### Módulo “start\_server.py”

La interfaz de usuario se hace accesible mediante un sencillo servicio de aplicaciones web desarrollado en Python, utilizando la librería “SimpleHTTPServer”.

Se accede al interfaz a través del acceso <http://localhost:8000>.

### Módulo “index.html”



Página web en formato HTML que contiene el contenido de la página de inicio con el menú en la parte superior y la estructura necesaria para albergar los componentes del interfaz de usuario.

Esta página web se ha diseñado como una aplicación de página única o *simple page application (SPA)* que se va renderizando con la lógica de la aplicación en el módulo *javascript* correspondiente.

### **Módulo “style.css”**

Hoja de estilos en formato *cascade style-sheet (css)* del interfaz de usuario web que contiene la configuración de los elementos del menú inicial principalmente.

### **Módulo “index.js”**

Fichero con el código *javascript* con la lógica del interfaz de usuario.

Este fichero contiene inicialmente todas las funciones que se ejecutan cuando se pulsan las diferentes opciones del menú, y son las que renderizan el interfaz de usuario correspondiente para cada funcionalidad del menú/submenú. Las funciones son: *inicio\_onclick()*, *new\_informes\_onclick()*, *query\_informes\_onclick()*, *new\_termin\_onclick()*, *query\_terminos\_onclick()*, *acerca\_onclick()*.

Posteriormente se ubican las funciones que se ejecutan cuando se desencadenan las diferentes acciones dentro de las funcionalidades del interfaz para invocar al motor de inferencia y renderizar los resultados obtenidos. Las funciones son: *callEngine\_Application()*, *callEngine\_newreport()*, *callEngine\_getterminosclinicos()*, *callEngine\_queryreports()*, *callEngine\_newtermin()*.

El código fuente está incluido en el repositorio de *github* correspondiente: <https://github.com/fmedinafernandez/14MBID-TFM-NEL>

## Apéndice V

### Validación de resultados DISTEMIST

A continuación, se presentan los datos obtenidos al validar los resultados por el prototipo con respecto a los datos etiquetados en DISTEMIST.

Identificador Informe	Correc- tos	Incorrec- tos	Faltan- tes	Etiqueta- dos	DISTE- MIST
es-S0004-06142007000600016-2	3	9	1	12	4
es-S0004-06142007000700013-1	8	6	4	14	12
es-S0004-06142007000700014-1	6	7	2	13	8
es-S0004-06142007000700015-1	2	7	6	9	8
es-S0004-06142007000900011-1	7	8	2	15	9
es-S0004-06142007000900013-1	1	0	0	1	1
es-S0004-06142007000900014-1	5	3	4	8	9
es-S0004-06142008000100008-1	3	10	2	13	5
es-S0004-06142008000100011-1	12	8	5	20	17
es-S0004-06142008000300015-1	4	5	0	9	4
es-S0004-06142008000500014-1	3	4	1	7	4
es-S0004-06142008000500015-1	4	4	2	8	6
es-S0004-06142008000600013-1	3	5	2	8	5
es-S0004-06142008000600014-1	5	15	3	20	8
es-S0004-06142008000700011-1	1	3	1	4	2
es-S0004-06142008000700015-2	4	5	0	9	4
es-S0004-06142009000100009-1	3	4	3	7	6
es-S0004-06142009000100010-1	10	11	2	21	12

es-S0004-06142009000100010-3	7	4	3	11	10
es-S0004-06142009000100013-1	3	13	7	16	10
es-S0004-06142009000200008-1	3	4	4	7	7
es-S0004-06142009000200013-1	4	8	3	12	7
es-S0004-06142009000300014-1	2	6	4	8	6
es-S0004-06142009000300014-4	2	10	4	12	6
es-S0004-06142009000500009-1	9	14	4	23	13
es-S0004-06142009000500011-1	8	16	4	24	12
es-S0004-06142009000500013-1	6	7	5	13	11
es-S0004-06142009000700011-1	7	16	4	23	11
es-S0004-06142009000800010-1	7	15	9	22	16
es-S0004-06142009000800011-1	1	8	1	9	2
es-S0004-06142009000900011-1	5	10	3	15	8
es-S0004-06142009000900012-1	13	14	8	27	21
es-S0004-06142010000300011-1	7	6	5	13	12
es-S0004-06142010000500011-1	8	16	4	24	12
es-S0004-06142010000500014-1	1	3	0	4	1
es-S0210-48062003000400012-1	5	2	2	7	7
es-S0210-48062003000900013-1	2	3	6	5	8
es-S0210-48062003001000009-1	15	11	16	26	31
es-S0210-48062004000300007-1	4	13	6	17	10
es-S0210-48062004000400008-1	9	21	7	30	16
es-S0210-48062004000500008-1	6	3	3	9	9
es-S0210-48062004000500010-	4	5	5	9	9

1					
es-S0210-48062004000500011-1	8	16	5	24	13
es-S0210-48062004000700006-1	9	4	0	13	9
es-S0210-48062005000100016-1	5	13	10	18	15
es-S0210-48062005000300016-2	16	13	7	29	23
es-S0210-48062005000600012-1	4	8	2	12	6
es-S0210-48062005000600013-1	3	12	5	15	8
es-S0210-48062005000700010-1	3	0	1	3	4
es-S0210-48062005000700013-1	2	2	4	4	6
es-S0210-48062005000700015-1	5	11	6	16	11
es-S0210-48062005000800014-1	7	12	4	19	11
es-S0210-48062005001000013-1	4	15	4	19	8
es-S0210-48062006000100012-1	5	4	2	9	7
es-S0210-48062006000100016-1	3	8	3	11	6
es-S0210-48062006000200017-1	7	14	4	21	11
es-S0210-48062006000300015-1	4	8	2	12	6
es-S0210-48062006000300016-1	14	18	5	32	19
es-S0210-48062006000700010-1	4	10	2	14	6
es-S0210-48062006000900012-1	4	6	0	10	4
es-S0210-48062007000100004-1	6	5	0	11	6
es-S0210-48062007000100011-1	6	9	2	15	8
es-S0210-48062007000100012-1	6	11	4	17	10
es-S0210-48062007000400013-1	3	19	3	22	6
es-S0210-48062007000700015-1	6	8	3	14	9

es-S0210-48062007000700016-1	1	7	4	8	5
es-S0210-48062007000700017-1	12	13	3	25	15
es-S0210-48062007001000016-1	14	15	4	29	18
es-S0210-48062009000200017-1	9	12	4	21	13
es-S0210-48062009000300013-2	2	6	5	8	7
es-S0210-48062009000300013-4	3	4	4	7	7
es-S0210-48062009000300013-6	5	6	2	11	7
es-S0210-48062009000400016-1	11	17	8	28	19
es-S0210-48062009000600016-1	3	8	3	11	6
es-S0210-48062009000600017-1	8	13	3	21	11
es-S0210-48062009000900015-1	6	7	3	13	9
es-S0210-48062009000900015-2	5	15	2	20	7
es-S0210-48062009000900015-3	5	7	0	12	5
es-S0210-48062009000900017-1	8	9	1	17	9
es-S0210-48062009000900019-1	5	3	4	8	9
es-S0210-48062009001000016-2	3	2	1	5	4
es-S0210-48062009001000019-1	4	6	6	10	10
es-S0210-48062010000100019-1	5	10	4	15	9
es-S0210-48062010000100019-3	2	7	3	9	5
es-S0210-56912006000200007-1	11	8	4	19	15
es-S0210-56912006000300007-2	5	8	3	13	8
es-S0210-56912006000600007-1	9	13	1	22	10
es-S0210-56912006000800008-1	9	12	3	21	12
es-S0210-56912007000200007-	8	3	2	11	10

3					
es-S0210-56912007000700008-1	3	4	1	7	4
es-S0210-56912007000900007-1	10	8	4	18	14
es-S0210-56912007000900007-3	6	4	1	10	7
es-S0210-56912007000900007-6	7	3	4	10	11
es-S0210-56912008000100008-1	3	3	2	6	5
es-S0210-56912008000100008-3	4	6	1	10	5
es-S0210-56912008000100008-4	4	3	1	7	5
es-S0210-56912008000200007-2	2	1	1	3	3
es-S0210-56912008000400007-1	5	7	0	12	5
es-S0210-56912008000400007-2	4	5	1	9	5
es-S0210-56912008000400007-4	3	4	1	7	4
es-S0210-56912009000600008-1	5	7	3	12	8
es-S0210-56912009000700006-2	5	9	2	14	7
es-S0210-56912009000900008-1	12	9	3	21	15
es-S0210-56912010000200009-1	2	4	2	6	4
es-S0210-56912011000500009-3	3	5	1	8	4
es-S0210-56912011000700007-1	11	10	2	21	13
es-S0210-56912012000900010-1	5	1	2	6	7
es-S0211-57352007000200017-1	9	4	5	13	14
es-S0211-57352011000100008-1	6	8	7	14	13
es-S0211-57352013000300012-1	19	5	10	24	29
es-S0211-69952009000500014-1	7	9	0	16	7
es-S0211-69952009000500017-1	6	5	4	11	10

es-S0211-69952010000400013-1	7	12	4	19	11
es-S0211-69952011000500011-1	6	7	3	13	9
es-S0211-69952011000500011-2	3	6	2	9	5
es-S0211-69952011000500011-3	5	5	1	10	6
es-S0211-69952011000500013-1	9	11	6	20	15
es-S0211-69952012000700031-1	8	8	2	16	10
es-S0211-69952013000200019-1	9	6	4	15	13
es-S0211-69952013000400027-1	3	6	1	9	4
es-S0211-69952013000500019-1	12	15	2	27	14
es-S0211-69952013000500032-2	11	6	2	17	13
es-S0211-69952013000500035-1	11	12	6	23	17
es-S0211-69952013000700023-1	3	8	2	11	5
es-S0211-69952013000700025-1	4	9	0	13	4
es-S0211-69952014000100016-1	11	21	5	32	16
es-S0211-69952014000200012-1	4	12	3	16	7
es-S0211-69952014000200012-2	3	6	5	9	8
es-S0211-69952014000200016-1	3	3	2	6	5
es-S0211-69952014000200018-1	13	14	2	27	15
es-S0211-69952014000400020-1	5	12	7	17	12
es-S0211-69952014000500020-1	7	4	1	11	8
es-S0211-69952015000100013-1	6	12	8	18	14
es-S0211-69952015000200015-1	12	13	4	25	16
es-S0211-69952015000300011-1	6	7	5	13	11
es-S0211-69952017000200225-	7	8	0	15	7

1					
es-S0211-69952017000200225-2	6	1	1	7	7
es-S0212-16112005000400011-1	5	5	3	10	8
es-S0212-16112005000600011-1	13	14	3	27	16
es-S0212-16112007000600015-1	6	6	4	12	10
es-S0212-16112007000700016-1	0	10	1	10	1
es-S0212-16112007000800012-1	10	13	8	23	18
es-S0212-16112009000300015-1	5	7	2	12	7
es-S0212-16112009000600017-1	4	10	3	14	7
es-S0212-16112011000600041-1	5	2	0	7	5
es-S0212-16112011000600041-2	9	5	3	14	12
es-S0212-16112012000300029-1	8	8	4	16	12
es-S0212-16112012000500042-1	8	11	1	19	9
es-S0212-16112012000500043-1	5	6	6	11	11
es-S0212-16112013000300055-1	13	15	4	28	17
es-S0212-16112013000300056-1	7	11	7	18	14
es-S0212-16112014000200032-3	3	6	0	9	3
es-S0212-16112014000600030-2	6	6	3	12	9
es-S0212-71992000001200010-1	4	7	10	11	14
es-S0212-71992003000500006-1	4	5	7	9	11
es-S0212-71992004000600005-1	12	20	8	32	20
es-S0212-71992004001100007-1	9	16	6	25	15
es-S0212-71992004001200007-1	7	18	3	25	10
es-S0212-71992005000700008-1	7	8	1	15	8



es-S0212-71992005001000009-1	2	5	2	7	4
es-S0212-71992005001200008-1	3	5	0	8	3
es-S0212-71992005001200010-1	13	17	7	30	20
es-S0212-71992006000100006-1	9	8	4	17	13
es-S0212-71992006000100008-1	9	12	5	21	14
es-S0212-71992006000200005-1	5	10	2	15	7
es-S0212-71992006000200006-1	12	16	6	28	18
es-S0212-71992006000600007-1	8	10	2	18	10
es-S0212-71992006000600008-1	9	5	4	14	13
es-S0212-71992006000800007-1	11	17	6	28	17
es-S0212-71992006000900007-1	12	21	6	33	18
es-S0212-71992006001000008-1	7	5	1	12	8
es-S0212-71992007000100006-1	11	21	5	32	16
es-S0212-71992007000100007-1	6	11	5	17	11
es-S0212-71992007000200007-1	8	5	3	13	11
es-S0212-71992007000300005-1	10	11	1	21	11
es-S0212-71992007000300006-1	8	6	5	14	13
es-S0212-71992007000400005-1	3	3	4	6	7
es-S0212-71992007000600008-1	9	18	7	27	16
es-S0212-71992007001000007-1	5	16	3	21	8
es-S0212-71992007001000008-1	3	4	2	7	5
es-S0213-12852003000600002-1	1	1	0	2	1
es-S0213-12852004000500002-1	2	3	1	5	3
es-S0213-12852005000200004-1	6	4	5	10	11

1					
es-S0213-12852016000400002-1	3	6	1	9	4
es-S0365-66912002001100008-1	4	6	4	10	8
es-S0365-66912003000600010-1	7	9	5	16	12
es-S0365-66912004000600008-1	3	10	2	13	5
es-S0365-66912004000600009-2	6	6	1	12	7
es-S0365-66912004001000009-1	6	14	5	20	11
es-S0365-66912004001100011-1	10	15	4	25	14
es-S0365-66912004001200011-1	7	18	4	25	11
es-S0365-66912005000600011-1	4	7	0	11	4
es-S0365-66912005001000011-1	8	9	3	17	11
es-S0365-66912005001200008-1	4	15	4	19	8
es-S0365-66912005001200010-1	16	19	3	35	19
es-S0365-66912005001200012-1	5	7	5	12	10
es-S0365-66912006000100009-1	4	9	0	13	4
es-S0365-66912006000300011-1	2	5	1	7	3
es-S0365-66912006000400010-1	9	20	5	29	14
es-S0365-66912006000500008-1	4	9	3	13	7
es-S0365-66912006000600010-1	2	8	1	10	3
es-S0365-66912006000600010-2	3	3	2	6	5
es-S0365-66912006000600011-1	9	7	3	16	12
es-S0365-66912006000800010-1	2	8	2	10	4
es-S0365-66912006000900011-1	11	9	5	20	16
es-S0365-66912006001000011-1	2	10	1	12	3

es-S0365-66912006001100007-1	5	6	3	11	8
es-S0365-66912006001200009-1	8	16	0	24	8
es-S0365-66912007000300010-1	8	10	1	18	9
es-S0365-66912007000700012-1	9	16	7	25	16
es-S0365-66912007000800012-1	7	8	3	15	10
es-S0365-66912007000800013-1	2	5	1	7	3
es-S0365-66912007000900014-1	6	13	5	19	11
es-S0365-66912007001100010-1	6	8	1	14	7
es-S0365-66912007001200012-1	4	10	3	14	7
es-S0365-66912007001200013-1	9	9	7	18	16
es-S0365-66912008000100007-1	6	6	3	12	9
es-S0365-66912008000100007-2	5	6	1	11	6
es-S0365-66912008000100009-3	3	8	7	11	10
es-S0365-66912008000200014-1	9	14	0	23	9
es-S0365-66912008000700008-1	8	14	3	22	11
es-S0365-66912008000700011-1	5	14	5	19	10
es-S0365-66912008001000010-1	3	11	3	14	6
es-S0365-66912009000100008-1	11	11	5	22	16
es-S0365-66912009000300010-1	7	13	3	20	10
es-S0365-66912009000800005-1	5	9	2	14	7
es-S0365-66912009000800005-2	2	6	2	8	4
es-S0365-66912009001100005-1	3	11	3	14	6
es-S0365-66912010000200005-1	6	4	0	10	6
es-S0365-66912010000300004-	3	14	5	17	8

1					
es-S0365-66912010000300005-1	6	7	7	13	13
es-S0365-66912010000600004-1	5	13	3	18	8
es-S0365-66912011000400005-1	3	8	7	11	10
es-S0365-66912011000400005-2	2	5	5	7	7
es-S0365-66912011000600005-2	4	9	1	13	5
es-S0365-66912011000600006-1	13	11	1	24	14
es-S0365-66912011000800006-1	6	9	2	15	8
es-S0365-66912011000900005-1	10	11	2	21	12
es-S0365-66912011001000003-1	1	0	0	1	1
es-S0365-66912011001000003-2	1	1	1	2	2
es-S0365-66912011001100006-1	4	11	0	15	4
es-S0365-66912012000200003-1	7	11	6	18	13
es-S0365-66912012000200005-1	7	18	5	25	12
es-S0365-66912012000500005-2	7	11	2	18	9
es-S0365-66912012001100004-1	3	15	7	18	10
es-S0376-78922007000400008-1	14	19	6	33	20
es-S0376-78922008000400007-1	5	5	4	10	9
es-S0376-78922009000200007-1	4	7	1	11	5
es-S0376-78922009000200008-1	7	2	0	9	7
es-S0376-78922009000200008-2	3	4	0	7	3
es-S0376-78922009000400002-1	3	5	4	8	7
es-S0376-78922011000200004-1	4	7	6	11	10
es-S0376-78922012000200008-1	4	5	2	9	6

es-S0376-78922012000300005-1	8	11	4	19	12
es-S0376-78922014000100013-1	3	9	4	12	7
es-S0376-78922014000200011-1	4	9	1	13	5
es-S0376-78922014000200012-1	7	8	4	15	11
es-S0376-78922014000300013-1	8	10	3	18	11
es-S0376-78922015000200009-1	6	11	10	17	16
es-S0376-78922015000200010-1	9	12	7	21	16
es-S0376-78922015000400013-1	4	8	0	12	4
es-S0376-78922016000300005-1	5	6	4	11	9
es-S0376-78922016000300010-1	5	14	7	19	12
es-S0376-78922017000100008-1	3	5	1	8	4
es-S0378-48352004000300006-1	9	12	4	21	13
es-S0378-48352005000100005-1	3	4	3	7	6
es-S0378-48352005000400004-1	3	5	4	8	7
es-S0378-48352005000500005-1	6	9	5	15	11
es-S0378-48352005000700007-2	4	6	4	10	8
es-S0378-48352005000900005-1	6	7	3	13	9
es-S0378-48352006000300006-1	9	12	3	21	12
es-S0378-48352006000300008-1	5	7	5	12	10
es-S0378-48352006000900004-1	6	3	2	9	8
es-S0465-546X2008000200010-1	6	1	0	7	6
es-S0465-546X2009000300008-1	3	6	14	9	17
es-S0465-546X2011000300007-1	7	6	5	13	12
es-S0465-546X2014000300010-1	4	3	4	7	8

1					
es-S0465-546X2014000400012-1	2	14	3	16	5
es-S1130-01082005000500013-1	7	8	2	15	9
es-S1130-01082005000800013-1	1	5	2	6	3
es-S1130-01082005001000011-1	3	6	3	9	6
es-S1130-01082005001000014-1	3	17	5	20	8
es-S1130-01082005001100014-1	7	10	6	17	13
es-S1130-01082005001200013-1	5	5	2	10	7
es-S1130-01082006000100014-1	8	14	5	22	13
es-S1130-01082006000700014-1	8	8	3	16	11
es-S1130-01082006000900015-1	6	10	5	16	11
es-S1130-01082006001000017-1	2	4	2	6	4
es-S1130-01082007000100014-1	8	9	7	17	15
es-S1130-01082007000300006-7	1	1	3	2	4
es-S1130-01082007000400015-1	6	9	0	15	6
es-S1130-01082007000500016-1	6	13	4	19	10
es-S1130-01082007000600012-1	7	5	5	12	12
es-S1130-01082007000700016-1	3	6	1	9	4
es-S1130-01082007001000017-1	4	8	0	12	4
es-S1130-01082007001100012-1	6	4	2	10	8
es-S1130-01082008000100009-1	8	7	2	15	10
es-S1130-01082008000100010-1	10	5	5	15	15
es-S1130-01082008000200009-1	2	1	2	3	4
es-S1130-01082008000200009-3	2	0	3	2	5

es-S1130-01082008000200015-1	0	8	7	8	7
es-S1130-01082008000400013-1	6	4	1	10	7
es-S1130-01082008000400017-1	6	16	10	22	16
es-S1130-01082008000500016-2	8	10	3	18	11
es-S1130-01082008000500017-2	2	5	3	7	5
es-S1130-01082008000800010-1	1	9	3	10	4
es-S1130-01082008000800011-1	11	12	5	23	16
es-S1130-01082008000800019-1	6	4	0	10	6
es-S1130-01082008000900014-1	5	5	1	10	6
es-S1130-01082008000900015-1	6	11	2	17	8
es-S1130-01082008000900016-1	4	10	1	14	5
es-S1130-01082008001100009-1	2	2	2	4	4
es-S1130-01082008001200016-1	2	3	2	5	4
es-S1130-01082008001200016-2	2	3	2	5	4
es-S1130-01082009000400009-1	13	13	3	26	16
es-S1130-01082009000400011-1	4	10	3	14	7
es-S1130-01082009000400014-1	4	5	1	9	5
es-S1130-01082009000500011-1	3	6	3	9	6
es-S1130-01082009000500015-1	9	9	1	18	10
es-S1130-01082009000600018-1	4	9	4	13	8
es-S1130-01082009000700012-1	4	7	2	11	6
es-S1130-01082009000700015-1	4	2	0	6	4
es-S1130-01082009000700016-1	2	6	2	8	4
es-S1130-01082009000700018-	3	6	2	9	5

1					
es-S1130-01082009000900015-1	6	6	5	12	11
es-S1130-01082009001000011-1	1	13	4	14	5
es-S1130-01082009001000013-1	9	12	5	21	14
es-S1130-01082010000100021-1	4	3	0	7	4
es-S1130-01082010000300015-1	4	7	4	11	8
es-S1130-01082010000600013-1	5	14	5	19	10
es-S1130-01082010000700016-1	13	12	0	25	13
es-S1130-05582003000600004-1	6	8	7	14	13
es-S1130-05582004000400006-1	6	17	10	23	16
es-S1130-05582004000500005-1	10	7	5	17	15
es-S1130-05582005000200003-1	4	4	2	8	6
es-S1130-05582007000500003-1	3	8	2	11	5
es-S1130-05582007000500006-1	3	4	4	7	7
es-S1130-05582007000600003-1	3	12	3	15	6
es-S1130-05582007000600006-1	4	8	2	12	6
es-S1130-05582008000300004-1	1	10	5	11	6
es-S1130-05582008000400007-2	2	2	2	4	4
es-S1130-05582008000400008-3	4	4	2	8	6
es-S1130-05582008000500006-1	9	7	1	16	10
es-S1130-05582008000500009-1	1	10	4	11	5
es-S1130-05582009000400004-1	2	4	3	6	5
es-S1130-05582009000400005-1	5	7	3	12	8
es-S1130-05582009000600003-3	2	2	0	4	2



es-S1130-05582009000600006-1	2	8	3	10	5
es-S1130-05582010000200005-1	2	15	3	17	5
es-S1130-05582011000200005-1	1	9	8	10	9
es-S1130-05582011000300005-1	9	9	1	18	10
es-S1130-05582012000100005-1	7	4	6	11	13
es-S1130-05582012000300006-1	1	7	7	8	8
es-S1130-05582013000200006-1	10	16	8	26	18
es-S1130-05582013000200007-1	1	10	4	11	5
es-S1130-05582013000400006-1	6	9	2	15	8
es-S1130-05582014000200008-1	9	13	13	22	22
es-S1130-05582014000400006-1	7	11	0	18	7
es-S1130-05582014000400008-1	6	7	4	13	10
es-S1130-05582015000100008-1	9	10	3	19	12
es-S1130-05582015000200006-1	1	4	1	5	2
es-S1130-05582015000300004-1	4	10	4	14	8
es-S1130-05582015000300007-1	4	14	1	18	5
es-S1130-05582015000400010-1	7	7	2	14	9
es-S1130-05582016000100005-1	5	9	1	14	6
es-S1130-05582016000400228-1	5	10	4	15	9
es-S1130-05582017000100031-1	1	0	2	1	3
es-S1130-05582017000100037-1	5	4	3	9	8
es-S1130-05582017000100044-2	8	6	4	14	12
es-S1130-05582017000200099-1	3	5	2	8	5
es-S1130-05582017000200105-	2	16	7	18	9

1					
es-S1130-05582017000200117-1	10	13	6	23	16
es-S1130-05582017000200122-1	8	18	5	26	13
es-S1130-05582017000300150-2	2	8	1	10	3
es-S1130-14732005000400006-1	7	16	1	23	8
es-S1130-14732005000400007-1	6	23	4	29	10
es-S1130-14732005000500006-1	7	9	2	16	9
es-S1130-14732005000600007-1	5	16	4	21	9
es-S1130-14732005000600008-1	8	4	4	12	12
es-S1130-14732006000400007-1	12	8	5	20	17
es-S1130-14732008000500005-1	7	5	2	12	9
es-S1130-14732008000500005-2	6	5	0	11	6
es-S1130-14732009000600006-1	1	7	8	8	9
es-S1130-14732011000300008-1	7	10	4	17	11
es-S1130-63432013000100011-1	6	19	6	25	12
es-S1130-63432013000300012-1	8	6	3	14	11
es-S1130-63432013000300013-1	5	9	2	14	7
es-S1130-63432013000400011-2	5	15	7	20	12
es-S1130-63432013000500013-1	1	4	6	5	7
es-S1130-63432013000500014-1	6	6	1	12	7
es-S1130-63432013000500017-1	5	5	1	10	6
es-S1130-63432013000600017-1	2	5	0	7	2
es-S1130-63432014000200011-1	9	5	3	14	12
es-S1130-63432014000300013-4	1	1	1	2	2

es-S1130-63432014000400015-1	2	4	1	6	3
es-S1130-63432014000500010-1	12	15	0	27	12
es-S1130-63432014000600009-1	10	8	6	18	16
es-S1130-63432015000500008-1	3	5	1	8	4
es-S1130-63432016000100007-1	2	2	1	4	3
es-S1130-63432016000100009-1	1	0	2	1	3
es-S1130-63432016000200009-1	2	1	0	3	2
es-S1130-63432016000400010-1	7	5	2	12	9
es-S1130-63432016000400011-1	9	11	5	20	14
es-S1130-63432016000500012-1	5	2	1	7	6
es-S1130-63432016000600013-1	1	0	0	1	1
es-S1131-57682001000100007-2	2	6	2	8	4
es-S1131-57682003000300008-1	9	8	5	17	14
es-S1131-57682003000400014-1	5	13	4	18	9
es-S1132-62552015000100006-1	4	10	5	14	9
es-S1134-80462004000400007-1	2	8	3	10	5
es-S1134-80462004000600007-1	4	6	4	10	8
es-S1134-80462005000300007-1	6	11	4	17	10
es-S1134-80462006000600005-1	6	13	4	19	10
es-S1134-80462008000200008-1	5	10	2	15	7
es-S1134-80462008000500004-2	1	0	0	1	1
es-S1134-80462008000800003-1	5	6	3	11	8
es-S1134-80462009000100005-15	0	3	1	3	1
es-S1134-80462009000100005-	0	5	1	5	1

16					
es-S1134-80462009000800004-1	3	7	2	10	5
es-S1134-80462009000800005-1	1	12	2	13	3
es-S1134-80462014000300004-1	2	6	0	8	2
es-S1134-80462014000300004-4	4	3	1	7	5
es-S1134-80462015000100006-1	14	13	0	27	14
es-S1134-80462015000200003-1	2	8	2	10	4
es-S1134-80462015000200004-1	17	9	4	26	21
es-S1134-80462015000300003-1	6	5	3	11	9
es-S1134-80462016000600005-3	4	5	7	9	11
es-S1135-76062003000400006-1	11	11	7	22	18
es-S1135-76062009000400004-1	5	6	1	11	6
es-S1135-76062010000200004-1	5	4	1	9	6
es-S1135-76062011000100006-1	7	5	2	12	9
es-S1135-76062011000200009-1	7	5	2	12	9
es-S1135-76062012000300014-1	8	8	6	16	14
es-S1135-76062014000200009-1	8	3	0	11	8
es-S1135-76062016000100006-1	5	6	5	11	10
es-S1137-66272004000500009-1	4	6	4	10	8
es-S1137-66272006000300009-1	8	8	3	16	11
es-S1137-66272006000500011-1	5	7	1	12	6
es-S1137-66272006000500012-1	8	7	0	15	8
es-S1137-66272008000300009-1	7	12	3	19	10
es-S1137-66272008000500009-1	5	11	6	16	11

es-S1137-66272009000100010-1	8	5	5	13	13
es-S1137-66272009000100013-1	9	6	3	15	12
es-S1137-66272009000300013-1	4	7	3	11	7
es-S1137-66272009000300015-1	1	24	5	25	6
es-S1137-66272009000500015-1	4	6	5	10	9
es-S1137-66272009000500017-1	3	6	2	9	5
es-S1137-66272010000300010-1	6	10	4	16	10
es-S1137-66272011000100012-1	8	4	6	12	14
es-S1137-66272011000100013-1	7	5	1	12	8
es-S1137-66272011000100013-3	6	3	0	9	6
es-S1137-66272011000300022-1	5	8	7	13	12
es-S1137-66272012000100011-1	10	12	8	22	18
es-S1137-66272012000100018-1	10	13	7	23	17
es-S1137-66272012000200017-1	3	2	1	5	4
es-S1137-66272012000300021-1	4	7	3	11	7
es-S1137-66272013000200019-1	6	1	6	7	12
es-S1137-66272013000200022-1	6	12	5	18	11
es-S1137-66272014000100021-1	5	9	3	14	8
es-S1137-66272014000300015-1	2	12	6	14	8
es-S1137-66272014000300016-1	7	16	5	23	12
es-S1137-66272014000300017-1	6	7	3	13	9
es-S1137-66272015000300014-1	7	11	5	18	12
es-S1137-66272016000200013-1	8	12	9	20	17
es-S1137-66272016000200015-1	5	11	4	16	9

1					
es-S1138-123X2003000200006-1	3	4	2	7	5
es-S1139-13752006000200009-1	4	9	9	13	13
es-S1139-13752009000200010-1	3	3	3	6	6
es-S1139-13752009000200010-2	6	3	1	9	7
es-S1139-76322009000200009-1	3	2	2	5	5
es-S1139-76322009000400007-1	2	8	4	10	6
es-S1139-76322009000500006-1	9	5	1	14	10
es-S1139-76322009000700014-1	12	11	7	23	19
es-S1139-76322009000700014-2	5	13	4	18	9
es-S1139-76322009000700015-1	8	14	6	22	14
es-S1139-76322010000100009-1	4	3	3	7	7
es-S1139-76322010000300008-1	3	3	2	6	5
es-S1139-76322010000600010-1	6	10	6	16	12
es-S1139-76322011000100011-1	4	4	3	8	7
es-S1139-76322011000200007-1	1	0	1	1	2
es-S1139-76322011000300007-2	1	5	0	6	1
es-S1139-76322011000300008-1	2	2	0	4	2
es-S1139-76322012000300006-2	2	5	2	7	4
es-S1139-76322012000400010-1	2	1	2	3	4
es-S1139-76322012000400011-1	2	1	1	3	3
es-S1139-76322013000100009-1	8	5	4	13	12
es-S1139-76322013000100010-1	1	3	5	4	6
es-S1139-76322013000400008-1	2	2	1	4	3

es-S1139-76322013000500010-1	5	5	1	10	6
es-S1139-76322014000100010-1	3	5	1	8	4
es-S1139-76322014000400007-1	5	9	4	14	9
es-S1139-76322014000400009-1	3	4	3	7	6
es-S1139-76322014000500011-1	8	11	3	19	11
es-S1139-76322014000500014-1	5	7	1	12	6
es-S1139-76322015000100010-1	4	10	3	14	7
es-S1139-76322015000100011-1	1	1	1	2	2
es-S1139-76322015000100012-3	1	1	0	2	1
es-S1139-76322015000100012-4	2	1	1	3	3
es-S1139-76322015000100013-1	4	11	1	15	5
es-S1139-76322015000100014-1	9	8	3	17	12
es-S1139-76322015000300015-1	2	1	1	3	3
es-S1139-76322015000300017-1	5	7	2	12	7
es-S1139-76322015000300018-1	6	4	4	10	10
es-S1139-76322015000300018-2	3	3	0	6	3
es-S1139-76322015000400006-1	3	3	1	6	4
es-S1139-76322015000400008-1	3	8	6	11	9
es-S1139-76322015000500009-1	1	2	1	3	2
es-S1139-76322016000200010-1	3	3	3	6	6
es-S1139-76322016000300007-1	1	0	0	1	1
es-S1139-76322016000300012-1	6	3	3	9	9
es-S1139-76322016000300013-1	2	3	1	5	3
es-S1139-76322016000400011-1	3	1	0	4	3

2					
es-S1139-76322016000400020-1	7	6	2	13	9
es-S1139-76322017000100011-1	3	10	0	13	3
es-S1139-76322017000200007-1	3	4	5	7	8
es-S1139-76322017000200009-1	7	11	7	18	14
es-S1139-76322017000200016-3	1	4	3	5	4
es-S1579-699X2004000400002-1	12	2	10	14	22
es-S1698-44472004000100009-1	5	9	2	14	7
es-S1698-44472004000200009-1	2	3	1	5	3
es-S1698-44472004000500007-1	4	9	6	13	10
es-S1698-44472005000100002-1	4	4	0	8	4
es-S1698-44472005000300015-1	3	3	2	6	5
es-S1698-44472005000400012-1	4	9	8	13	12
es-S1698-69462006000100017-1	5	10	2	15	7
es-S1698-69462006000100019-1	8	10	3	18	11
es-S1698-69462006000300015-1	2	4	2	6	4
es-S1698-69462006000400002-1	4	10	4	14	8
es-S1699-65852010000300002-1	2	6	4	8	6
es-S1699-695X2014000200007-1	6	8	3	14	9
es-S1699-695X2014000200012-1	6	15	7	21	13
es-S1699-695X2014000300006-1	9	13	10	22	19
es-S1699-695X2014000300011-1	9	13	1	22	10
es-S1699-695X2015000100013-1	9	4	2	13	11
es-S1699-695X2015000200013-1	8	7	2	15	10



es-S1699-695X2016000200009-1	9	16	6	25	15
es-S1699-695X2016000200010-1	6	15	1	21	7
es-S1699-695X2017000100011-1	19	8	1	27	20
es-S1887-85712011000200007-1	3	9	1	12	4
es-S1887-85712013000200013-1	1	6	2	7	3
es-S1887-85712015000100005-1	8	17	8	25	16
es-S1887-85712015000200004-1	3	11	4	14	7
es-S1887-85712015000200005-1	2	1	1	3	3
es-S1887-85712017000100004-1	5	5	3	10	8
es-S1888-75462017000100042-1	2	8	1	10	3
es-S1889-836X2015000100003-1	11	14	9	25	20
es-S1889-836X2016000100006-1	15	8	4	23	19
es-S2254-28842013000300009-1	14	9	9	23	23
es-S2254-28842014000300010-1	6	14	4	20	10
S0004-06142005000700014-1	5	6	2	11	7
S0004-06142005000900013-1	3	3	0	6	3
S0004-06142005000900014-1	6	5	3	11	9
S0004-06142005000900015-1	5	3	0	8	5
S0004-06142005000900016-1	9	2	1	11	10
S0004-06142005001000011-1	10	18	3	28	13
S0004-06142005001000011-3	10	12	1	22	11
S0004-06142005001000016-1	3	2	3	5	6
S0004-06142006000200001-1	4	3	3	7	7
S0004-06142006000200011-1	2	9	2	11	4
S0004-06142006000200013-1	5	7	2	12	7
S0004-06142006000200014-1	2	7	5	9	7
S0004-06142006000300015-1	3	14	6	17	9
S0004-06142006000500002-4	4	5	2	9	6
S0004-06142006000500011-1	9	17	10	26	19
S0004-06142006000600009-1	5	10	5	15	10
S0004-06142006000600012-1	5	16	12	21	17
S0004-06142006000600014-1	2	1	1	3	3

S0004-06142006000600015-1	4	3	2	7	6
S0004-06142006000700008-1	9	14	7	23	16
S0004-06142006000700012-1	5	5	7	10	12
S0004-06142006000700013-1	7	10	7	17	14
S0004-06142006000900006-1	2	4	3	6	5
S0004-06142006000900008-1	7	2	3	9	10
S0004-06142007000100011-1	10	14	4	24	14
S0004-06142007000100015-1	10	11	2	21	12
S0004-06142007000200011-1	8	11	4	19	12
S0004-06142007000300011-1	8	8	8	16	16
S0004-06142007000300012-1	4	12	3	16	7
S0004-06142007000300013-1	4	2	1	6	5
S0004-06142007000500011-1	2	6	1	8	3
S0004-06142007000500014-1	4	5	1	9	5
S0004-06142007000500017-1	1	2	0	3	1
S0004-06142007000600011-1	1	2	1	3	2

Tabla 17. Tabla con el resultado al validar los datos etiquetados por el prototipo en relación con DISTEMIST. Elaboración propia.

**Identificador de informe:** Esta columna muestra el identificador del informe y nombre del fichero que contiene el informe en DISTEMIST.

**Correctos:** Esta columna muestra el número de etiquetas correctamente vinculadas en relación con los datos de DISTEMIST.

**Incorrectos:** Esta columna muestra el número de etiquetas incorrectamente vinculadas en relación con los datos de DISTEMIST, es decir, que son etiquetas que no pertenecen al informe.

**Faltantes:** Esta columna muestra el número de etiquetas no vinculadas en relación con los datos de DISTEMIST, es decir, que son etiquetas que no se han detectado en el informe.

Identificador Informe	%Correctos sobre DISTEMIST	%Faltantes sobre DISTEMIST	%Incorrectos sobre etiquetados
es-S0004-06142007000600016-2	75%	25%	75%
es-S0004-06142007000700013-1	67%	33%	43%
es-S0004-06142007000700014-1	75%	25%	54%
es-S0004-06142007000700015-1	25%	75%	78%
es-S0004-06142007000900011-	78%	22%	53%

1			
es-S0004-06142007000900013-1	100%	0%	0%
es-S0004-06142007000900014-1	56%	44%	38%
es-S0004-06142008000100008-1	60%	40%	77%
es-S0004-06142008000100011-1	71%	29%	40%
es-S0004-06142008000300015-1	100%	0%	56%
es-S0004-06142008000500014-1	75%	25%	57%
es-S0004-06142008000500015-1	67%	33%	50%
es-S0004-06142008000600013-1	60%	40%	63%
es-S0004-06142008000600014-1	63%	38%	75%
es-S0004-06142008000700011-1	50%	50%	75%
es-S0004-06142008000700015-2	100%	0%	56%
es-S0004-06142009000100009-1	50%	50%	57%
es-S0004-06142009000100010-1	83%	17%	52%
es-S0004-06142009000100010-3	70%	30%	36%
es-S0004-06142009000100013-1	30%	70%	81%
es-S0004-06142009000200008-1	43%	57%	57%
es-S0004-06142009000200013-1	57%	43%	67%
es-S0004-06142009000300014-1	33%	67%	75%
es-S0004-06142009000300014-4	33%	67%	83%
es-S0004-06142009000500009-1	69%	31%	61%
es-S0004-06142009000500011-1	67%	33%	67%
es-S0004-06142009000500013-1	55%	45%	54%
es-S0004-06142009000700011-1	64%	36%	70%

es-S0004-06142009000800010-1	44%	56%	68%
es-S0004-06142009000800011-1	50%	50%	89%
es-S0004-06142009000900011-1	63%	38%	67%
es-S0004-06142009000900012-1	62%	38%	52%
es-S0004-06142010000300011-1	58%	42%	46%
es-S0004-06142010000500011-1	67%	33%	67%
es-S0004-06142010000500014-1	100%	0%	75%
es-S0210-48062003000400012-1	71%	29%	29%
es-S0210-48062003000900013-1	25%	75%	60%
es-S0210-48062003001000009-1	48%	52%	42%
es-S0210-48062004000300007-1	40%	60%	76%
es-S0210-48062004000400008-1	56%	44%	70%
es-S0210-48062004000500008-1	67%	33%	33%
es-S0210-48062004000500010-1	44%	56%	56%
es-S0210-48062004000500011-1	62%	38%	67%
es-S0210-48062004000700006-1	100%	0%	31%
es-S0210-48062005000100016-1	33%	67%	72%
es-S0210-48062005000300016-2	70%	30%	45%
es-S0210-48062005000600012-1	67%	33%	67%
es-S0210-48062005000600013-1	38%	63%	80%
es-S0210-48062005000700010-1	75%	25%	0%
es-S0210-48062005000700013-1	33%	67%	50%
es-S0210-48062005000700015-1	45%	55%	69%
es-S0210-48062005000800014-	64%	36%	63%

1			
es-S0210-48062005001000013-1	50%	50%	79%
es-S0210-48062006000100012-1	71%	29%	44%
es-S0210-48062006000100016-1	50%	50%	73%
es-S0210-48062006000200017-1	64%	36%	67%
es-S0210-48062006000300015-1	67%	33%	67%
es-S0210-48062006000300016-1	74%	26%	56%
es-S0210-48062006000700010-1	67%	33%	71%
es-S0210-48062006000900012-1	100%	0%	60%
es-S0210-48062007000100004-1	100%	0%	45%
es-S0210-48062007000100011-1	75%	25%	60%
es-S0210-48062007000100012-1	60%	40%	65%
es-S0210-48062007000400013-1	50%	50%	86%
es-S0210-48062007000700015-1	67%	33%	57%
es-S0210-48062007000700016-1	20%	80%	88%
es-S0210-48062007000700017-1	80%	20%	52%
es-S0210-48062007001000016-1	78%	22%	52%
es-S0210-48062009000200017-1	69%	31%	57%
es-S0210-48062009000300013-2	29%	71%	75%
es-S0210-48062009000300013-4	43%	57%	57%
es-S0210-48062009000300013-6	71%	29%	55%
es-S0210-48062009000400016-1	58%	42%	61%
es-S0210-48062009000600016-1	50%	50%	73%
es-S0210-48062009000600017-1	73%	27%	62%

es-S0210-48062009000900015-1	67%	33%	54%
es-S0210-48062009000900015-2	71%	29%	75%
es-S0210-48062009000900015-3	100%	0%	58%
es-S0210-48062009000900017-1	89%	11%	53%
es-S0210-48062009000900019-1	56%	44%	38%
es-S0210-48062009001000016-2	75%	25%	40%
es-S0210-48062009001000019-1	40%	60%	60%
es-S0210-48062010000100019-1	56%	44%	67%
es-S0210-48062010000100019-3	40%	60%	78%
es-S0210-56912006000200007-1	73%	27%	42%
es-S0210-56912006000300007-2	63%	38%	62%
es-S0210-56912006000600007-1	90%	10%	59%
es-S0210-56912006000800008-1	75%	25%	57%
es-S0210-56912007000200007-3	80%	20%	27%
es-S0210-56912007000700008-1	75%	25%	57%
es-S0210-56912007000900007-1	71%	29%	44%
es-S0210-56912007000900007-3	86%	14%	40%
es-S0210-56912007000900007-6	64%	36%	30%
es-S0210-56912008000100008-1	60%	40%	50%
es-S0210-56912008000100008-3	80%	20%	60%
es-S0210-56912008000100008-4	80%	20%	43%
es-S0210-56912008000200007-2	67%	33%	33%
es-S0210-56912008000400007-1	100%	0%	58%
es-S0210-56912008000400007-	80%	20%	56%

2			
es-S0210-56912008000400007-4	75%	25%	57%
es-S0210-56912009000600008-1	63%	38%	58%
es-S0210-56912009000700006-2	71%	29%	64%
es-S0210-56912009000900008-1	80%	20%	43%
es-S0210-56912010000200009-1	50%	50%	67%
es-S0210-56912011000500009-3	75%	25%	63%
es-S0210-56912011000700007-1	85%	15%	48%
es-S0210-56912012000900010-1	71%	29%	17%
es-S0211-57352007000200017-1	64%	36%	31%
es-S0211-57352011000100008-1	46%	54%	57%
es-S0211-57352013000300012-1	66%	34%	21%
es-S0211-69952009000500014-1	100%	0%	56%
es-S0211-69952009000500017-1	60%	40%	45%
es-S0211-69952010000400013-1	64%	36%	63%
es-S0211-69952011000500011-1	67%	33%	54%
es-S0211-69952011000500011-2	60%	40%	67%
es-S0211-69952011000500011-3	83%	17%	50%
es-S0211-69952011000500013-1	60%	40%	55%
es-S0211-69952012000700031-1	80%	20%	50%
es-S0211-69952013000200019-1	69%	31%	40%
es-S0211-69952013000400027-1	75%	25%	67%
es-S0211-69952013000500019-1	86%	14%	56%
es-S0211-69952013000500032-2	85%	15%	35%

es-S0211-69952013000500035-1	65%	35%	52%
es-S0211-69952013000700023-1	60%	40%	73%
es-S0211-69952013000700025-1	100%	0%	69%
es-S0211-69952014000100016-1	69%	31%	66%
es-S0211-69952014000200012-1	57%	43%	75%
es-S0211-69952014000200012-2	38%	63%	67%
es-S0211-69952014000200016-1	60%	40%	50%
es-S0211-69952014000200018-1	87%	13%	52%
es-S0211-69952014000400020-1	42%	58%	71%
es-S0211-69952014000500020-1	88%	13%	36%
es-S0211-69952015000100013-1	43%	57%	67%
es-S0211-69952015000200015-1	75%	25%	52%
es-S0211-69952015000300011-1	55%	45%	54%
es-S0211-69952017000200225-1	100%	0%	53%
es-S0211-69952017000200225-2	86%	14%	14%
es-S0212-16112005000400011-1	63%	38%	50%
es-S0212-16112005000600011-1	81%	19%	52%
es-S0212-16112007000600015-1	60%	40%	50%
es-S0212-16112007000700016-1	0%	100%	100%
es-S0212-16112007000800012-1	56%	44%	57%
es-S0212-16112009000300015-1	71%	29%	58%
es-S0212-16112009000600017-1	57%	43%	71%
es-S0212-16112011000600041-1	100%	0%	29%
es-S0212-16112011000600041-	75%	25%	36%



2			
es-S0212-16112012000300029-1	67%	33%	50%
es-S0212-16112012000500042-1	89%	11%	58%
es-S0212-16112012000500043-1	45%	55%	55%
es-S0212-16112013000300055-1	76%	24%	54%
es-S0212-16112013000300056-1	50%	50%	61%
es-S0212-16112014000200032-3	100%	0%	67%
es-S0212-16112014000600030-2	67%	33%	50%
es-S0212-71992000001200010-1	29%	71%	64%
es-S0212-71992003000500006-1	36%	64%	56%
es-S0212-71992004000600005-1	60%	40%	63%
es-S0212-71992004001100007-1	60%	40%	64%
es-S0212-71992004001200007-1	70%	30%	72%
es-S0212-71992005000700008-1	88%	13%	53%
es-S0212-71992005001000009-1	50%	50%	71%
es-S0212-71992005001200008-1	100%	0%	63%
es-S0212-71992005001200010-1	65%	35%	57%
es-S0212-71992006000100006-1	69%	31%	47%
es-S0212-71992006000100008-1	64%	36%	57%
es-S0212-71992006000200005-1	71%	29%	67%
es-S0212-71992006000200006-1	67%	33%	57%
es-S0212-71992006000600007-1	80%	20%	56%
es-S0212-71992006000600008-1	69%	31%	36%
es-S0212-71992006000800007-1	65%	35%	61%

es-S0212-71992006000900007-1	67%	33%	64%
es-S0212-71992006001000008-1	88%	13%	42%
es-S0212-71992007000100006-1	69%	31%	66%
es-S0212-71992007000100007-1	55%	45%	65%
es-S0212-71992007000200007-1	73%	27%	38%
es-S0212-71992007000300005-1	91%	9%	52%
es-S0212-71992007000300006-1	62%	38%	43%
es-S0212-71992007000400005-1	43%	57%	50%
es-S0212-71992007000600008-1	56%	44%	67%
es-S0212-71992007001000007-1	63%	38%	76%
es-S0212-71992007001000008-1	60%	40%	57%
es-S0213-12852003000600002-1	100%	0%	50%
es-S0213-12852004000500002-1	67%	33%	60%
es-S0213-12852005000200004-1	55%	45%	40%
es-S0213-12852016000400002-1	75%	25%	67%
es-S0365-66912002001100008-1	50%	50%	60%
es-S0365-66912003000600010-1	58%	42%	56%
es-S0365-66912004000600008-1	60%	40%	77%
es-S0365-66912004000600009-2	86%	14%	50%
es-S0365-66912004001000009-1	55%	45%	70%
es-S0365-66912004001100011-1	71%	29%	60%
es-S0365-66912004001200011-1	64%	36%	72%
es-S0365-66912005000600011-1	100%	0%	64%
es-S0365-66912005001000011-1	73%	27%	53%

1			
es-S0365-66912005001200008-1	50%	50%	79%
es-S0365-66912005001200010-1	84%	16%	54%
es-S0365-66912005001200012-1	50%	50%	58%
es-S0365-66912006000100009-1	100%	0%	69%
es-S0365-66912006000300011-1	67%	33%	71%
es-S0365-66912006000400010-1	64%	36%	69%
es-S0365-66912006000500008-1	57%	43%	69%
es-S0365-66912006000600010-1	67%	33%	80%
es-S0365-66912006000600010-2	60%	40%	50%
es-S0365-66912006000600011-1	75%	25%	44%
es-S0365-66912006000800010-1	50%	50%	80%
es-S0365-66912006000900011-1	69%	31%	45%
es-S0365-66912006001000011-1	67%	33%	83%
es-S0365-66912006001100007-1	63%	38%	55%
es-S0365-66912006001200009-1	100%	0%	67%
es-S0365-66912007000300010-1	89%	11%	56%
es-S0365-66912007000700012-1	56%	44%	64%
es-S0365-66912007000800012-1	70%	30%	53%
es-S0365-66912007000800013-1	67%	33%	71%
es-S0365-66912007000900014-1	55%	45%	68%
es-S0365-66912007001100010-1	86%	14%	57%
es-S0365-66912007001200012-1	57%	43%	71%
es-S0365-66912007001200013-1	56%	44%	50%

es-S0365-66912008000100007-1	67%	33%	50%
es-S0365-66912008000100007-2	83%	17%	55%
es-S0365-66912008000100009-3	30%	70%	73%
es-S0365-66912008000200014-1	100%	0%	61%
es-S0365-66912008000700008-1	73%	27%	64%
es-S0365-66912008000700011-1	50%	50%	74%
es-S0365-66912008001000010-1	50%	50%	79%
es-S0365-66912009000100008-1	69%	31%	50%
es-S0365-66912009000300010-1	70%	30%	65%
es-S0365-66912009000800005-1	71%	29%	64%
es-S0365-66912009000800005-2	50%	50%	75%
es-S0365-66912009001100005-1	50%	50%	79%
es-S0365-66912010000200005-1	100%	0%	40%
es-S0365-66912010000300004-1	38%	63%	82%
es-S0365-66912010000300005-1	46%	54%	54%
es-S0365-66912010000600004-1	63%	38%	72%
es-S0365-66912011000400005-1	30%	70%	73%
es-S0365-66912011000400005-2	29%	71%	71%
es-S0365-66912011000600005-2	80%	20%	69%
es-S0365-66912011000600006-1	93%	7%	46%
es-S0365-66912011000800006-1	75%	25%	60%
es-S0365-66912011000900005-1	83%	17%	52%
es-S0365-66912011001000003-1	100%	0%	0%
es-S0365-66912011001000003-	50%	50%	50%

2			
es-S0365-66912011001100006-1	100%	0%	73%
es-S0365-66912012000200003-1	54%	46%	61%
es-S0365-66912012000200005-1	58%	42%	72%
es-S0365-66912012000500005-2	78%	22%	61%
es-S0365-66912012001100004-1	30%	70%	83%
es-S0376-78922007000400008-1	70%	30%	58%
es-S0376-78922008000400007-1	56%	44%	50%
es-S0376-78922009000200007-1	80%	20%	64%
es-S0376-78922009000200008-1	100%	0%	22%
es-S0376-78922009000200008-2	100%	0%	57%
es-S0376-78922009000400002-1	43%	57%	63%
es-S0376-78922011000200004-1	40%	60%	64%
es-S0376-78922012000200008-1	67%	33%	56%
es-S0376-78922012000300005-1	67%	33%	58%
es-S0376-78922014000100013-1	43%	57%	75%
es-S0376-78922014000200011-1	80%	20%	69%
es-S0376-78922014000200012-1	64%	36%	53%
es-S0376-78922014000300013-1	73%	27%	56%
es-S0376-78922015000200009-1	38%	63%	65%
es-S0376-78922015000200010-1	56%	44%	57%
es-S0376-78922015000400013-1	100%	0%	67%
es-S0376-78922016000300005-1	56%	44%	55%
es-S0376-78922016000300010-1	42%	58%	74%

es-S0376-78922017000100008-1	75%	25%	63%
es-S0378-48352004000300006-1	69%	31%	57%
es-S0378-48352005000100005-1	50%	50%	57%
es-S0378-48352005000400004-1	43%	57%	63%
es-S0378-48352005000500005-1	55%	45%	60%
es-S0378-48352005000700007-2	50%	50%	60%
es-S0378-48352005000900005-1	67%	33%	54%
es-S0378-48352006000300006-1	75%	25%	57%
es-S0378-48352006000300008-1	50%	50%	58%
es-S0378-48352006000900004-1	75%	25%	33%
es-S0465-546X2008000200010-1	100%	0%	14%
es-S0465-546X2009000300008-1	18%	82%	67%
es-S0465-546X2011000300007-1	58%	42%	46%
es-S0465-546X2014000300010-1	50%	50%	43%
es-S0465-546X2014000400012-1	40%	60%	88%
es-S1130-01082005000500013-1	78%	22%	53%
es-S1130-01082005000800013-1	33%	67%	83%
es-S1130-01082005001000011-1	50%	50%	67%
es-S1130-01082005001000014-1	38%	63%	85%
es-S1130-01082005001100014-1	54%	46%	59%
es-S1130-01082005001200013-1	71%	29%	50%
es-S1130-01082006000100014-1	62%	38%	64%
es-S1130-01082006000700014-1	73%	27%	50%
es-S1130-01082006000900015-	55%	45%	63%

1			
es-S1130-01082006001000017-1	50%	50%	67%
es-S1130-01082007000100014-1	53%	47%	53%
es-S1130-01082007000300006-7	25%	75%	50%
es-S1130-01082007000400015-1	100%	0%	60%
es-S1130-01082007000500016-1	60%	40%	68%
es-S1130-01082007000600012-1	58%	42%	42%
es-S1130-01082007000700016-1	75%	25%	67%
es-S1130-01082007001000017-1	100%	0%	67%
es-S1130-01082007001100012-1	75%	25%	40%
es-S1130-01082008000100009-1	80%	20%	47%
es-S1130-01082008000100010-1	67%	33%	33%
es-S1130-01082008000200009-1	50%	50%	33%
es-S1130-01082008000200009-3	40%	60%	0%
es-S1130-01082008000200015-1	0%	100%	100%
es-S1130-01082008000400013-1	86%	14%	40%
es-S1130-01082008000400017-1	38%	63%	73%
es-S1130-01082008000500016-2	73%	27%	56%
es-S1130-01082008000500017-2	40%	60%	71%
es-S1130-01082008000800010-1	25%	75%	90%
es-S1130-01082008000800011-1	69%	31%	52%
es-S1130-01082008000800019-1	100%	0%	40%
es-S1130-01082008000900014-1	83%	17%	50%
es-S1130-01082008000900015-1	75%	25%	65%

es-S1130-01082008000900016-1	80%	20%	71%
es-S1130-01082008001100009-1	50%	50%	50%
es-S1130-01082008001200016-1	50%	50%	60%
es-S1130-01082008001200016-2	50%	50%	60%
es-S1130-01082009000400009-1	81%	19%	50%
es-S1130-01082009000400011-1	57%	43%	71%
es-S1130-01082009000400014-1	80%	20%	56%
es-S1130-01082009000500011-1	50%	50%	67%
es-S1130-01082009000500015-1	90%	10%	50%
es-S1130-01082009000600018-1	50%	50%	69%
es-S1130-01082009000700012-1	67%	33%	64%
es-S1130-01082009000700015-1	100%	0%	33%
es-S1130-01082009000700016-1	50%	50%	75%
es-S1130-01082009000700018-1	60%	40%	67%
es-S1130-01082009000900015-1	55%	45%	50%
es-S1130-01082009001000011-1	20%	80%	93%
es-S1130-01082009001000013-1	64%	36%	57%
es-S1130-01082010000100021-1	100%	0%	43%
es-S1130-01082010000300015-1	50%	50%	64%
es-S1130-01082010000600013-1	50%	50%	74%
es-S1130-01082010000700016-1	100%	0%	48%
es-S1130-05582003000600004-1	46%	54%	57%
es-S1130-05582004000400006-1	38%	63%	74%
es-S1130-05582004000500005-1	67%	33%	41%



1			
es-S1130-05582005000200003-1	67%	33%	50%
es-S1130-05582007000500003-1	60%	40%	73%
es-S1130-05582007000500006-1	43%	57%	57%
es-S1130-05582007000600003-1	50%	50%	80%
es-S1130-05582007000600006-1	67%	33%	67%
es-S1130-05582008000300004-1	17%	83%	91%
es-S1130-05582008000400007-2	50%	50%	50%
es-S1130-05582008000400008-3	67%	33%	50%
es-S1130-05582008000500006-1	90%	10%	44%
es-S1130-05582008000500009-1	20%	80%	91%
es-S1130-05582009000400004-1	40%	60%	67%
es-S1130-05582009000400005-1	63%	38%	58%
es-S1130-05582009000600003-3	100%	0%	50%
es-S1130-05582009000600006-1	40%	60%	80%
es-S1130-05582010000200005-1	40%	60%	88%
es-S1130-05582011000200005-1	11%	89%	90%
es-S1130-05582011000300005-1	90%	10%	50%
es-S1130-05582012000100005-1	54%	46%	36%
es-S1130-05582012000300006-1	13%	88%	88%
es-S1130-05582013000200006-1	56%	44%	62%
es-S1130-05582013000200007-1	20%	80%	91%
es-S1130-05582013000400006-1	75%	25%	60%
es-S1130-05582014000200008-1	41%	59%	59%

es-S1130-05582014000400006-1	100%	0%	61%
es-S1130-05582014000400008-1	60%	40%	54%
es-S1130-05582015000100008-1	75%	25%	53%
es-S1130-05582015000200006-1	50%	50%	80%
es-S1130-05582015000300004-1	50%	50%	71%
es-S1130-05582015000300007-1	80%	20%	78%
es-S1130-05582015000400010-1	78%	22%	50%
es-S1130-05582016000100005-1	83%	17%	64%
es-S1130-05582016000400228-1	56%	44%	67%
es-S1130-05582017000100031-1	33%	67%	0%
es-S1130-05582017000100037-1	63%	38%	44%
es-S1130-05582017000100044-2	67%	33%	43%
es-S1130-05582017000200099-1	60%	40%	63%
es-S1130-05582017000200105-1	22%	78%	89%
es-S1130-05582017000200117-1	63%	38%	57%
es-S1130-05582017000200122-1	62%	38%	69%
es-S1130-05582017000300150-2	67%	33%	80%
es-S1130-14732005000400006-1	88%	13%	70%
es-S1130-14732005000400007-1	60%	40%	79%
es-S1130-14732005000500006-1	78%	22%	56%
es-S1130-14732005000600007-1	56%	44%	76%
es-S1130-14732005000600008-1	67%	33%	33%
es-S1130-14732006000400007-1	71%	29%	40%
es-S1130-14732008000500005-	78%	22%	42%

1			
es-S1130-14732008000500005-2	100%	0%	45%
es-S1130-14732009000600006-1	11%	89%	88%
es-S1130-14732011000300008-1	64%	36%	59%
es-S1130-63432013000100011-1	50%	50%	76%
es-S1130-63432013000300012-1	73%	27%	43%
es-S1130-63432013000300013-1	71%	29%	64%
es-S1130-63432013000400011-2	42%	58%	75%
es-S1130-63432013000500013-1	14%	86%	80%
es-S1130-63432013000500014-1	86%	14%	50%
es-S1130-63432013000500017-1	83%	17%	50%
es-S1130-63432013000600017-1	100%	0%	71%
es-S1130-63432014000200011-1	75%	25%	36%
es-S1130-63432014000300013-4	50%	50%	50%
es-S1130-63432014000400015-1	67%	33%	67%
es-S1130-63432014000500010-1	100%	0%	56%
es-S1130-63432014000600009-1	63%	38%	44%
es-S1130-63432015000500008-1	75%	25%	63%
es-S1130-63432016000100007-1	67%	33%	50%
es-S1130-63432016000100009-1	33%	67%	0%
es-S1130-63432016000200009-1	100%	0%	33%
es-S1130-63432016000400010-1	78%	22%	42%
es-S1130-63432016000400011-1	64%	36%	55%
es-S1130-63432016000500012-1	83%	17%	29%

es-S1130-63432016000600013-1	100%	0%	0%
es-S1131-57682001000100007-2	50%	50%	75%
es-S1131-57682003000300008-1	64%	36%	47%
es-S1131-57682003000400014-1	56%	44%	72%
es-S1132-62552015000100006-1	44%	56%	71%
es-S1134-80462004000400007-1	40%	60%	80%
es-S1134-80462004000600007-1	50%	50%	60%
es-S1134-80462005000300007-1	60%	40%	65%
es-S1134-80462006000600005-1	60%	40%	68%
es-S1134-80462008000200008-1	71%	29%	67%
es-S1134-80462008000500004-2	100%	0%	0%
es-S1134-80462008000800003-1	63%	38%	55%
es-S1134-80462009000100005-15	0%	100%	100%
es-S1134-80462009000100005-16	0%	100%	100%
es-S1134-80462009000800004-1	60%	40%	70%
es-S1134-80462009000800005-1	33%	67%	92%
es-S1134-80462014000300004-1	100%	0%	75%
es-S1134-80462014000300004-4	80%	20%	43%
es-S1134-80462015000100006-1	100%	0%	48%
es-S1134-80462015000200003-1	50%	50%	80%
es-S1134-80462015000200004-1	81%	19%	35%
es-S1134-80462015000300003-1	67%	33%	45%
es-S1134-80462016000600005-3	36%	64%	56%
es-S1135-76062003000400006-	61%	39%	50%

1			
es-S1135-76062009000400004-1	83%	17%	55%
es-S1135-76062010000200004-1	83%	17%	44%
es-S1135-76062011000100006-1	78%	22%	42%
es-S1135-76062011000200009-1	78%	22%	42%
es-S1135-76062012000300014-1	57%	43%	50%
es-S1135-76062014000200009-1	100%	0%	27%
es-S1135-76062016000100006-1	50%	50%	55%
es-S1137-66272004000500009-1	50%	50%	60%
es-S1137-66272006000300009-1	73%	27%	50%
es-S1137-66272006000500011-1	83%	17%	58%
es-S1137-66272006000500012-1	100%	0%	47%
es-S1137-66272008000300009-1	70%	30%	63%
es-S1137-66272008000500009-1	45%	55%	69%
es-S1137-66272009000100010-1	62%	38%	38%
es-S1137-66272009000100013-1	75%	25%	40%
es-S1137-66272009000300013-1	57%	43%	64%
es-S1137-66272009000300015-1	17%	83%	96%
es-S1137-66272009000500015-1	44%	56%	60%
es-S1137-66272009000500017-1	60%	40%	67%
es-S1137-66272010000300010-1	60%	40%	63%
es-S1137-66272011000100012-1	57%	43%	33%
es-S1137-66272011000100013-1	88%	13%	42%
es-S1137-66272011000100013-3	100%	0%	33%

es-S1137-66272011000300022-1	42%	58%	62%
es-S1137-66272012000100011-1	56%	44%	55%
es-S1137-66272012000100018-1	59%	41%	57%
es-S1137-66272012000200017-1	75%	25%	40%
es-S1137-66272012000300021-1	57%	43%	64%
es-S1137-66272013000200019-1	50%	50%	14%
es-S1137-66272013000200022-1	55%	45%	67%
es-S1137-66272014000100021-1	63%	38%	64%
es-S1137-66272014000300015-1	25%	75%	86%
es-S1137-66272014000300016-1	58%	42%	70%
es-S1137-66272014000300017-1	67%	33%	54%
es-S1137-66272015000300014-1	58%	42%	61%
es-S1137-66272016000200013-1	47%	53%	60%
es-S1137-66272016000200015-1	56%	44%	69%
es-S1138-123X2003000200006-1	60%	40%	57%
es-S1139-13752006000200009-1	31%	69%	69%
es-S1139-13752009000200010-1	50%	50%	50%
es-S1139-13752009000200010-2	86%	14%	33%
es-S1139-76322009000200009-1	60%	40%	40%
es-S1139-76322009000400007-1	33%	67%	80%
es-S1139-76322009000500006-1	90%	10%	36%
es-S1139-76322009000700014-1	63%	37%	48%
es-S1139-76322009000700014-2	56%	44%	72%
es-S1139-76322009000700015-	57%	43%	64%

1			
es-S1139-76322010000100009-1	57%	43%	43%
es-S1139-76322010000300008-1	60%	40%	50%
es-S1139-76322010000600010-1	50%	50%	63%
es-S1139-76322011000100011-1	57%	43%	50%
es-S1139-76322011000200007-1	50%	50%	0%
es-S1139-76322011000300007-2	100%	0%	83%
es-S1139-76322011000300008-1	100%	0%	50%
es-S1139-76322012000300006-2	50%	50%	71%
es-S1139-76322012000400010-1	50%	50%	33%
es-S1139-76322012000400011-1	67%	33%	33%
es-S1139-76322013000100009-1	67%	33%	38%
es-S1139-76322013000100010-1	17%	83%	75%
es-S1139-76322013000400008-1	67%	33%	50%
es-S1139-76322013000500010-1	83%	17%	50%
es-S1139-76322014000100010-1	75%	25%	63%
es-S1139-76322014000400007-1	56%	44%	64%
es-S1139-76322014000400009-1	50%	50%	57%
es-S1139-76322014000500011-1	73%	27%	58%
es-S1139-76322014000500014-1	83%	17%	58%
es-S1139-76322015000100010-1	57%	43%	71%
es-S1139-76322015000100011-1	50%	50%	50%
es-S1139-76322015000100012-3	100%	0%	50%
es-S1139-76322015000100012-4	67%	33%	33%

es-S1139-76322015000100013-1	80%	20%	73%
es-S1139-76322015000100014-1	75%	25%	47%
es-S1139-76322015000300015-1	67%	33%	33%
es-S1139-76322015000300017-1	71%	29%	58%
es-S1139-76322015000300018-1	60%	40%	40%
es-S1139-76322015000300018-2	100%	0%	50%
es-S1139-76322015000400006-1	75%	25%	50%
es-S1139-76322015000400008-1	33%	67%	73%
es-S1139-76322015000500009-1	50%	50%	67%
es-S1139-76322016000200010-1	50%	50%	50%
es-S1139-76322016000300007-1	100%	0%	0%
es-S1139-76322016000300012-1	67%	33%	33%
es-S1139-76322016000300013-1	67%	33%	60%
es-S1139-76322016000400011-2	100%	0%	25%
es-S1139-76322016000400020-1	78%	22%	46%
es-S1139-76322017000100011-1	100%	0%	77%
es-S1139-76322017000200007-1	38%	63%	57%
es-S1139-76322017000200009-1	50%	50%	61%
es-S1139-76322017000200016-3	25%	75%	80%
es-S1579-699X2004000400002-1	55%	45%	14%
es-S1698-44472004000100009-1	71%	29%	64%
es-S1698-44472004000200009-1	67%	33%	60%
es-S1698-44472004000500007-1	40%	60%	69%
es-S1698-44472005000100002-	100%	0%	50%



1			
es-S1698-44472005000300015-1	60%	40%	50%
es-S1698-44472005000400012-1	33%	67%	69%
es-S1698-69462006000100017-1	71%	29%	67%
es-S1698-69462006000100019-1	73%	27%	56%
es-S1698-69462006000300015-1	50%	50%	67%
es-S1698-69462006000400002-1	50%	50%	71%
es-S1699-65852010000300002-1	33%	67%	75%
es-S1699-695X2014000200007-1	67%	33%	57%
es-S1699-695X2014000200012-1	46%	54%	71%
es-S1699-695X2014000300006-1	47%	53%	59%
es-S1699-695X2014000300011-1	90%	10%	59%
es-S1699-695X2015000100013-1	82%	18%	31%
es-S1699-695X2015000200013-1	80%	20%	47%
es-S1699-695X2016000200009-1	60%	40%	64%
es-S1699-695X2016000200010-1	86%	14%	71%
es-S1699-695X2017000100011-1	95%	5%	30%
es-S1887-85712011000200007-1	75%	25%	75%
es-S1887-85712013000200013-1	33%	67%	86%
es-S1887-85712015000100005-1	50%	50%	68%
es-S1887-85712015000200004-1	43%	57%	79%
es-S1887-85712015000200005-1	67%	33%	33%
es-S1887-85712017000100004-1	63%	38%	50%
es-S1888-75462017000100042-1	67%	33%	80%

es-S1889-836X2015000100003-1	55%	45%	56%
es-S1889-836X2016000100006-1	79%	21%	35%
es-S2254-28842013000300009-1	61%	39%	39%
es-S2254-28842014000300010-1	60%	40%	70%
S0004-06142005000700014-1	71%	29%	55%
S0004-06142005000900013-1	100%	0%	50%
S0004-06142005000900014-1	67%	33%	45%
S0004-06142005000900015-1	100%	0%	38%
S0004-06142005000900016-1	90%	10%	18%
S0004-06142005001000011-1	77%	23%	64%
S0004-06142005001000011-3	91%	9%	55%
S0004-06142005001000016-1	50%	50%	40%
S0004-06142006000200001-1	57%	43%	43%
S0004-06142006000200011-1	50%	50%	82%
S0004-06142006000200013-1	71%	29%	58%
S0004-06142006000200014-1	29%	71%	78%
S0004-06142006000300015-1	33%	67%	82%
S0004-06142006000500002-4	67%	33%	56%
S0004-06142006000500011-1	47%	53%	65%
S0004-06142006000600009-1	50%	50%	67%
S0004-06142006000600012-1	29%	71%	76%
S0004-06142006000600014-1	67%	33%	33%
S0004-06142006000600015-1	67%	33%	43%
S0004-06142006000700008-1	56%	44%	61%
S0004-06142006000700012-1	42%	58%	50%
S0004-06142006000700013-1	50%	50%	59%
S0004-06142006000900006-1	40%	60%	67%
S0004-06142006000900008-1	70%	30%	22%
S0004-06142007000100011-1	71%	29%	58%
S0004-06142007000100015-1	83%	17%	52%
S0004-06142007000200011-1	67%	33%	58%
S0004-06142007000300011-1	50%	50%	50%
S0004-06142007000300012-1	57%	43%	75%
S0004-06142007000300013-1	80%	20%	33%
S0004-06142007000500011-1	67%	33%	75%
S0004-06142007000500014-1	80%	20%	56%
S0004-06142007000500017-1	100%	0%	67%
S0004-06142007000600011-1	50%	50%	67%

Tabla 18. Tabla con el resultado al validar los datos etiquetados por el prototipo en relación con DISTEMIST. Elaboración propia.

**Identificador de informe:** Esta columna muestra el identificador del informe y nombre del fichero que contiene el informe en DISTEMIST.

**%Correctos sobre Distemist:** Esta columna muestra el porcentaje de etiquetas correctamente vinculadas con respecto al total de etiquetas de DISTEMIST.

**%Faltantes sobre Distemist:** Esta columna muestra el porcentaje de etiquetas que no han sido vinculadas con respecto al total de etiquetas de DISTEMIST.

**%Incorrectos sobre etiquetados:** Esta columna muestra el porcentaje de etiquetas vinculadas incorrectamente con respecto al total de etiquetas vinculadas por el prototipo.