

# IXA-NER-RE en eHealth-KD Challenge 2020: Aprendizaje por transferencia translingüe para medicina Extracción de relaciones

Edgar Andrésa , Óscar Sainz , Aitziber Atutxaa y Oier López de Lacallea

a Grupo IXA PNL, Universidad del País Vasco (UPV/EHU);

## Resumen El

eHealth-KD 2020 planteó este año un desafío de extracción automática de una amplia gama de conocimientos a partir de documentos de salud escritos en lengua española. Nuestro grupo ha participado en todos los escenarios propuestos; la principal, la subtask Named Entity Recognition (NER), la subtask Relation Extraction (RE), y el dominio alternativo obteniendo resultados muy diferentes en cada una de ellas. La tarea principal ha sido concebida como una tubería de las subtasks NER y RE, cada una de ellas desarrollada independientemente de la otra. La tarea de reconocimiento de entidades de nombres se ha concebido como un sistema seq2seq básico que aplica un modelo de lenguaje de propósito general e incrustaciones estáticas. A diferencia de la subtask NER, en la subtask ER se exploraron con éxito varios enfoques; En primer lugar, transferir métodos de aprendizaje como una forma de medir la capacidad de adaptación de modelos de lenguaje previamente entrenados tanto al dominio médico como al idioma español. En segundo lugar, Emparejar los espacios en blanco para abordar el problema del tamaño reducido del corpus de entrenamiento al producir representaciones de relaciones directamente a partir de texto no etiquetado. Como se mencionó, los resultados en las diferentes tareas fueron heterogéneos; mientras que el resultado en NER está en la media (F1 0,66), con amplio margen de mejora, el resultado en RE ha sido sobresaliente, obteniendo el primer puesto en esta tarea (F1 0,633) con más de 3 puntos sobre los siguientes clasificados. , demostrando la solidez de las técnicas propuestas.

## Palabras clave

Modelos de lenguaje, coincidencia de espacios en blanco, reconocimiento de entidades nombradas, extracción de relaciones

## 1. Introducción

En este artículo describimos nuestra participación en la tarea compartida eHealth-KD 2020 [1], consistente en extraer información semántica estructurada para textos médicos en español. El desafío se divide en dos tareas principales propuestas como tubería. La primera tarea está dedicada a la identificación y clasificación de entidades médicas. En la segunda tarea, los participantes deben detectar las relaciones semánticas entre las entidades, presumiblemente descubiertas en la primera tarea.

Los organizadores propusieron diferentes esquemas de evaluación en los que 1) los sistemas se evalúan en todas las tareas a la vez (evaluación principal), y 2) el reconocimiento de entidades y la extracción de relaciones se evalúan por separado (tarea A y tarea B, respectivamente). Nuestro sistema está construido sobre dos componentes independientes y, por lo tanto, la capacitación y el desarrollo de cada componente se lleva a cabo en sus subtasks específicas por separado.

Actas del Foro de Evaluación de Lenguas Ibéricas (IberLEF 2020) correo electrónico:

eandres011@ehu.eus (E. Andrés); osainz006@ehu.eus (O. Sainz); aitziber.attuxa@ehu.eus (A. Atutxa); oier.lopezofthetstreet@ehu.eus (VIEJO Lacalle)

orcid:



© 2020 Copyright de este artículo por parte de sus autores. Uso permitido bajo la Licencia Creative Commons Attribution 4.0 International (CC BY 4.0).



ISSN 1613-0073 Actas del taller CEUR (CEUR-WS.org)

Nos acercamos a la tarea de Reconocimiento de entidades nombradas (NER)<sup>1</sup> con un etiquetador de secuencia BiLSTM basado en caracteres [2] entrenado en el conjunto de entrenamiento proporcionado por los organizadores que emplean tanto un modelo de lenguaje de dominio general previamente entrenado como incrustaciones de palabras estáticas. Con respecto a la tarea de extracción de relaciones<sup>2</sup>, para resolverla, decidimos utilizar una estrategia de aprendizaje por transferencia y ajustar los modelos de lenguaje multilingües previamente entrenados existentes [3] en los datos anotados de la tarea.

Tenga en cuenta que proponemos un sistema con componentes heterogéneos, con diferentes objetivos para cada uno de la pieza en el sistema. De esta manera, los objetivos de nuestra participación en la tarea son dobles.

- Reconocimiento de entidades: nuestro objetivo era comprobar la idoneidad de un sistema basado en un modelo de lenguaje previamente entrenado basado en caracteres en un entorno NER heterogéneo. Los sistemas basados en modelos de lenguaje previamente entrenados se han utilizado con éxito en tareas de reconocimiento de entidades médicas (MER). Pero a diferencia de otros desafíos similares que involucraron MER (CLEF eHealth 2020, PharmacoNER 2019), la tarea actual es especialmente desafiante porque las entidades no son puramente médicas sino muy heterogéneas no solo semánticamente en cuanto al dominio, sino también sintácticamente.
- Extracción de relaciones: Nuestro principal objetivo al utilizar grandes modelos de lenguaje multilingües previamente entrenados es medir la capacidad de adaptación al dominio médico y al idioma español cuando se utilizan métodos de aprendizaje por transferencia. Además, realice experimentos adaptando Matching the Blanks [4] (MTB) a la configuración de eHealth-KD 2020.

El sistema obtiene resultados muy descompensados, mientras que en relación extracción superamos con creces al resto de participantes (3,4 puntos mejor que el segundo clasificado), el sistema tiene un gran margen de mejora en el reconocimiento de entidades (aún estamos más de 10 puntos por debajo de mejores sistemas). En general, nuestro sistema muestra resultados muy competitivos con 55,7 de F1 en la prueba principal.

## 2. Trabajo relacionado

El reconocimiento de entidades MER, a diferencia de NER, muestra ciertas especificidades [5], como su naturaleza descriptiva, su productividad y el uso masivo de siglas. Estas especificidades y el hecho de que los sistemas NER emplearon sistemáticamente incorporaciones estáticas, llevan a los investigadores a utilizar corpus en el dominio, en lugar de corpus de dominio general, tanto para entrenar los sistemas MER como las incorporaciones estáticas previamente entrenadas, con el fin de para obtener mejores resultados ya que controlar el dominio conduce a un mejor control de la polisemia ([6],[7]). Recientemente, el desempeño de las tareas NER y MER ha mostrado un avance significativo con la introducción de incrustaciones de palabras contextualizadas (ELMo [8], ULMFiT [9], BERT [10] y FLAIR [2]).

Aunque las incrustaciones contextualizadas parecen reducir la brecha entre el corpus general y el corpus específico de dominio, varios trabajos sobre tareas MER sostienen que las incrustaciones contextualizadas de dominio específico todavía producen un rendimiento superior sobre las incrustaciones de palabras estándar y de dominio general ([11],[12],[13],[14]). Como se mencionó en la introducción, la tarea actual del MER debido a su heterogeneidad (los conceptos son más específicos del dominio médico mientras que las acciones o referencias

---

<sup>1</sup>Las entidades se clasifican en 4 tipos: concepto, acción, predicado y referencia.

<sup>2</sup>Los 13 tipos de relaciones están organizados en 4 categorías principales: relaciones generales, relaciones contextuales, roles de acción y roles de predicado.

son menos específicos) representa un campo de juego perfecto para comprobar el rendimiento de las incrustaciones basadas en modelos de lenguaje contextualizados calculados sobre el corpus de dominio general en las diferentes entidades.

**Transferencia de aprendizaje** Recientemente, la transferencia de aprendizaje se ha demostrado como una alternativa exitosa cuando (casi) no hay datos anotados disponibles en el dominio y el idioma de destino [10, 4]. Los modelos recientes de secuencia de transformadores [15] superan el estado del arte en muchas tareas de extracción de información, como la extracción de relaciones [4, 16, 17]. Algunos trabajos intentan integrar la información disponible en las bases de conocimiento en modelos de secuencia de Transformers [16]. Sin embargo, enfoques más simples basados en marcadores de entidad (más detalles en la Sección 4) muestran el mismo desempeño competitivo con una configuración más rápida [4]. De manera similar, los modelos de lenguajes multilingües [3] han demostrado una capacidad impresionante para realizar un aprendizaje cero en tareas multilingües. Este tipo de modelos parece muy prometedor para tareas de extracción de relaciones donde el idioma de destino contiene pequeña formación anotada.

colocar.

**Aumento de datos** Se ha propuesto una variedad de aumento de datos para tareas de extracción de información. Uno de los paradigmas más importantes es la supervisión a distancia [18, 19], en la que las relaciones existentes en las bases de conocimiento se alinean con texto sin etiquetar basándose en algunas heurísticas y etiquetando automáticamente los datos de entrenamiento [20]. Más recientemente, Soares et al. [4] introducen un método de aumento que no requiere etiqueta de relación y adapta el aprendizaje del modelo mediante Matching the Blanks (MTB). En este trabajo exploramos la idea de MTB para abordar la tarea de extracción de relaciones en eHealth-KD 2020.

### 3. Sistema de reconocimiento de entidades

Adoptamos un enfoque de secuencia a secuencia de aprendizaje profundo [2] para llevar a cabo la tarea de reconocimiento de entidades nombradas (NER).

#### 3.1. Arquitectura NER

El sistema FLAIR [2] empleado para NER se muestra en la Figura 1 y está compuesto por tres componentes principales ; Primero, un modelo de lenguaje basado en caracteres (LM) que genera representaciones de palabras contextuales basadas en caracteres muy poderosas , que luego se concatenan con incrustaciones estáticas. En la parte superior de esta capa LM, una capa BiLSTM captura las dependencias secuenciales entre las palabras de la secuencia de entrada. Y finalmente, una capa de campos aleatorios condicionales (CRF) para manejar la inferencia de etiquetado.

El LM en este caso concatena: incrustaciones estáticas e incrustaciones de estilo contextual. Las incrustaciones contextuales de FLAIR se forman a partir de cálculos parciales basados en caracteres; la estrategia BiLSTM se utiliza para tener en cuenta el contexto. Esos cálculos se realizan como se ve en la parte inferior de la Figura 1, los resultados se concatenan con incrustaciones estáticas.

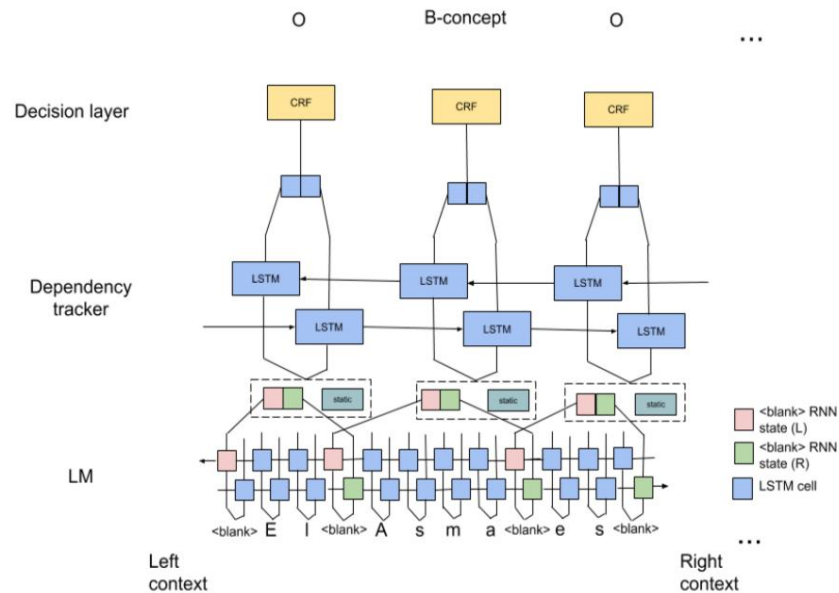


Figura 1: Arquitectura general del sistema NER.

La entrada proporcionada por el organizador como formato de enfrentamiento BRAT se tokenizó utilizando la función de propósito general `word_tokenize` de NLTK y luego se convirtió al formato Inside Outside Be-ginning (IOB). Este formato no captura entidades superpuestas y separadas. El conjunto de desarrollo se dividió de modo que una parte se utilizó para el desarrollo y otra para las pruebas. La salida del sistema se convirtió al formato de separación BRAT requerido en los archivos correspondientes (.ann), que consisten en una entidad, compensaciones y el texto coincidente.

### 3.2. Configuración de aprendizaje

La arquitectura propuesta se presentó en el enfoque actual, el modelo de lenguaje estaba compuesto por incrustaciones de palabras estáticas y contextuales, entre el LM y la capa de dependencia se propuso una capa de abandono, finalmente se conectó la capa de predicción después de ella. La arquitectura utilizada para el entrenamiento se puede ver en la Tabla 1. Los hiperparámetros de entrenamiento se pueden resumir de la siguiente manera: tasa de aprendizaje 0,1, tamaño de lote 16 y paciencia 3 para una parada temprana; tiene en cuenta el sobreajuste en el archivo de desarrollo. Se realizaron un máximo de cien épocas de entrenamiento y se detuvieron en 81 épocas. El proceso se calculó en el procesador CPU AMD Ryzen 7 1700 de ocho núcleos y tardó 45 minutos en finalizar.

En el experimento actual utilizamos incrustaciones estáticas de FastText previamente entrenadas (es-crawl) [21] entrenadas en rastros web (dominio general) e incrustaciones de estilo contextual (es-forward + es-backward) [2] entrenadas con Wikipedia (general). -dominio). Todas las capas de incrustación se calcularon manteniendo los parámetros predeterminados. No utilizamos las frases adicionales de Medline para entrenar al LM. Por lo tanto, no se llevó a cabo ningún ajuste en el dominio.

Capa	Abandonar	Especificación
LM-Adelante	0,5 - -	Incrustación (275, 100) LSTM (100, 2048) Lineal (entrada = 2048, salida = 275, sesgo = Verdadero)
LM-hacia atrás	0,5 - - -	Incrustación (275, 100) LSTM (100, 2048) Lineal (entrada = 2048, salida = 275, sesgo = Verdadero) Incrustación (275, 100)
Rastreador de dependencia	0,5+0,05(palabra) -	Lineal (entrada = 4396, salida = 4396, sesgo = Verdadero) BiLSTM (entrada = 4396, salida = 600)
Capa de decisión	-	Lineal (entrada = 600, salida = 11, sesgo = Verdadero, crf = verdadero)

tabla 1

Configuración del hiperparámetro NER.

#### 4. Sistema de extracción de relaciones.

En esta sección describimos nuestro componente de extracción de relaciones (RE). En total hemos construido tres Sistemas RE: XLMem, XLMem\* y XLMem\*+MTB. Todos los modelos se basan en el mismo XLM. con arquitectura de marcadores de entidad (XLMem), pero difieren en las estrategias de entrenamiento y los datos. Nosotros Primero describa la arquitectura base de los modelos XLMem. En las siguientes secciones, discutimos diferentes estrategias de entrenamiento y los valores de hiperparámetros utilizados en el entrenamiento.

##### 4.1. Arquitectura XLMem

La arquitectura básica de nuestro sistema es el codificador de relaciones. El codificador consta de un transformador. [15] modelo de lenguaje previamente entrenado basado en un cabezal de extracción de relaciones en la parte superior. Una particularidad de este codificador de relaciones es la necesidad de marcadores de entidad [4] como tokens adicionales en la entrada oración. Estos tokens especiales delimitan los límites de cada entidad en la oración de entrada como como se muestra en la Figura 2. La entrada consciente de la entidad luego se envía al modelo de lenguaje previamente entrenado. El cabezal de extracción de relaciones concatena las representaciones de los marcadores que indican la posición inicial de las entidades y combinarlas con una capa lineal que codifica la relación final representación.

Formalmente, dada una declaración de relación = (, 1, 2) formada por una secuencia de tokens = [ 0, 1, ..., ] y dos entidades y 2, primero corrompimos nuestra oración de entrada agregando la entidad marcadores ([E1S] y [E1E] definen dónde comienza y termina la primera entidad )

$$\sim = [ 0, \dots, [ 1 ], 1, [ 1 ], \dots, [ 2 ], 2, [ 2 ], \dots, ]$$

luego obtenemos las representaciones ocultas = relación  $\tilde{r}$  y finalmente definimos nuestra codificador de la siguiente manera:

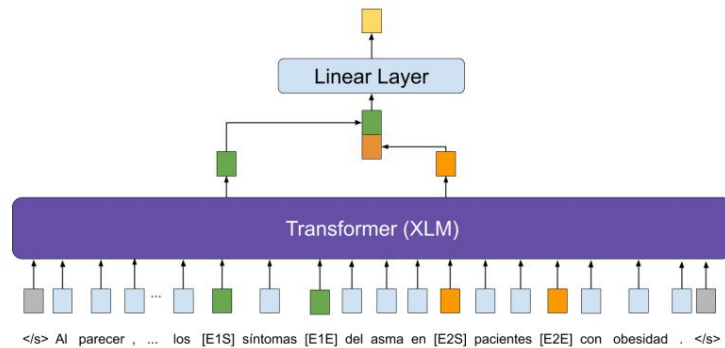


Figura 2: Arquitectura del codificador de relaciones XLMem basada en la estrategia Entity Marker para la representación de relaciones.

$$= [1; 2] + \quad (1)$$

dónde siendo  $2 \times y$  el tamaño de representación oculta. Finalmente, clasificación se realiza apilando una capa lineal encima del codificador con una activación softmax función:

$$() = ( ) + \quad (2)$$

donde  $\times y$  siendo el tamaño de representación oculta y el número de las relaciones.

El uso de XLM como modelo de lenguaje previamente entrenado brinda la oportunidad de aprender un lenguaje multilingüe. codificador de relación, que parece ser una buena opción para esta configuración. Concretamente, utilizamos el punto de control xlm-mlm-17-1280 proporcionado por el equipo de Hugging Face [22]. Este punto de control en particular ha sido capacitado en la estrategia Masked Language Model (MLM) con 17 idiomas, incluidos Español, que es nuestro idioma de destino para la tarea.

#### 4.2. Emparejar los espacios en blanco

Matching the Blanks [4] (MTB) puede considerarse una alternativa novedosa al conocido Distant Supervisión [18]. El enfoque se basa en la hipótesis de que si dos entidades están relacionadas, Es más probable que las oraciones que contienen esas dos entidades expresen la misma relación. La figura 3 muestra tres oraciones diferentes de nuestro corpus MTB. Mientras que las dos primeras oraciones codifican la misma relación para paciente y síntomas, la tercera oración expresa la relación para paciente and tiempo.

El conjunto de datos de entrenamiento se genera de la siguiente manera. Generamos pares positivos de oraciones (es decir, ejemplos 1 y 2 en la Figura 3) si comparten entidades en blanco. Generamos pares negativos fuertes si comparten una entidad (es decir, ejemplos 1 y 3), y negativos débiles si no se comparte ninguna entidad. Una vez que tengamos generado esos ejemplos, entrenamos un modelo que aprende si un par de oraciones codifica la misma relación o no, y transferimos los parámetros aprendidos a la extracción de la relación real tarea. Tenga en cuenta que se introducen [en blanco] para evitar simplemente volver a aprender y vincular entidades al conocimiento. Base (KB) utilizada para generar el corpus MTB.

- (1) Se observó actividad de CK en [blank] con dengue con presencia de [blank] como vómito, hematemesis y dolor abdominal.
- (2) Al parecer, existen mecanismos comunes a ambas patologías que pueden influir en la exacerbación de los [blank] del asma en [blank] con obesidad.
- (3) El [blank] promedio para el inicio de ENT fue de 30 (23,5) horas, y el 88,7% de los [blank] alcanzaron el objetivo nutricional en 48 horas.

Figura 3: Tres entradas diferentes en el conjunto de datos MTB. Los dos primeros comparten las mismas entidades paciente y síntomas y el tercero que contiene las entidades paciente y tiempo comparte solo una de ellas.

Hiperparámetro	Preentrenamiento MTB	sintonía FINA
Tasa de aprendizaje	1 -4	3 -4
Optimizador	SGD	SGD
Tamaño del lote	8	decadito
Pasos de acumulación de gradiente	8	4
Precisión de punto flotante	FP16 y FP32 FP16 y FP32	
Parada temprana (paciencia 3)		
Probabilidad de enmascaramiento de espacios en blanco	0,7	-

Tabla 2  
Configuración de hiperparámetros del proceso de capacitación del sistema de extracción de relaciones.

Para construir el corpus MTB hemos utilizado resúmenes de Medline en español. Los hemos procesado con Freeling [23] para extraer entidades médicas. En total conseguimos 7, 543 entidades médicas que conforma 278, 956 pares de entidades. Con esos pares de entidades hemos generado 691, 392 instancias positivas y 833, 332 casos negativos. Hemos dividido nuestros datos en 80% para entrenamiento y 20% para desarrollo. Finalmente, por razones técnicas descartamos aquellas instancias que contienen contextos. más de 128 tokens.

4.3. Configuración de aprendizaje

En esta sección informamos el conjunto de hiperparámetros que hemos utilizado en el momento del ajuste fino. los modelos (Figura 2). Para nuestro caso los hiperparámetros que mejor se ajustan al desarrollo Los conjuntos fueron los mismos para los tres enfoques probados. También informamos los hiperparámetros utilizados. durante el preentrenamiento de MTB.

Las configuraciones informadas se utilizan en una única GPU NVIDIA Titan V con 12 Gb de RAM. El proceso de puesta a punto tarda menos de 10 horas y en el caso del preentrenamiento de MTB tenemos detenido manualmente por razones de tiempo y plazos.

5. Resultados

La Tabla 3 muestra los resultados para las tareas de dominio principal y alternativo. Nuestra carrera oficial combina el sistema NER con el sistema XLMem RE. En este caso XLMem utilice el adicional

Modelo	Evaluación principal.			Alternativa		
	Prec.	Rec.	F1	Prec.	Rec.	F1
Vicomtech 0,679 Talp-UPC		0,652	0,665 0,594	0,626	0,535	0,563
0,626 UH-MAJA-KD	0,634	0,626	0,604 0,625	0,608	0,563	0,583
		0,615			0,498	0,547
IXA-DOWN-RE	0,536	0,580	0,557 0,563		0,416	0,478

Tabla 3

Resultados oficiales de los mejores sistemas en las tareas del dominio principal y alternativo.

Modelo	Rec. de desarrollo.			Prueba		
	Prec.	Rec.	F1	Prec.	Rec.	F1
SINAÍ	-	-	-	0,845 0,807	0,825	
vicomtec	-	-	-	0,822 0,820	0,821	
Talp-UPC	-	-	-	0,807 0,825	0,816	
UH-MAJA-KD	-	-	-	0,820 0,808	0,814	
Línea base de	-	-	-	0,795 0,825	0,767	
UH-MatCom	-	-	-	0,542 0,504	0,586	
(Nuestro) BiLSTM + CRF 0,742 0,696 0,718 0,692 0,727 0,660						

Tabla 4

Resultados de la tarea de reconocimiento de entidades nombradas. El sistema BiLSTM + CRF es el que se ha enviado a la competencia.

3.000 frases anotadas automáticamente de Medline que se proporcionaron para formación adicional. Los resultados generales muestran que, aunque nuestro sistema es competitivo (cuarto puesto general), todavía tiene grandes margen de mejora en la tarea principal, así como en la tarea alternativa. Nos gustaría  
Tenga en cuenta que una evaluación interna adicional mostró que nuestra mejor combinación de sistemas sería utilizar XLMem sin utilizar datos anotados automáticos adicionales.

### 5.1. Tarea de extracción de entidades (A)

En la siguiente Tabla 4 se pueden ver los resultados del Test de la tarea NER, los mejores resultados para cada Las métricas están resaltadas con caracteres en negrita, también proporcionamos nuestros resultados sobre el conjunto de desarrollo y sobre el conjunto de prueba oficial.

Aunque lejos del resultado obtenido por los primeros clasificados, el sistema presentado supera la línea de base sin ajustes finos y usando incorporaciones estáticas de dominio general y pre-entrenados Modelo de lenguaje. Un análisis de error preliminar nos ha llevado a concluir que, contrariamente a lo que pensamos Inicialmente pensé que el dominio podría ser relevante en esta tarea NER. Aunque 3 de los cuatro tipos de entidades (acciones, referencias y predicados) no son tipos de entidades de dominio especialmente médico, el hecho de que predecir referencias y predicados está fuertemente condicionado a haber predijo correctamente su concepto antecedente, y este último es la mayoría de las veces de dominio específico



Modelo	Tren			desarrollo			Prueba		
	Prec.	Rec.	F1	Prec.	Rec.	F1	Prec.	Rec.	F1
vicomtec	-	-	-	-	-	-	0,672	0,515	0,583
UH-MAJA-KD	-	-	-	-	-	-	0,629	0,571	0,599
(Nuestro) XLMem	0,861	0,849	0,855	0,708	0,642	0,674	0,690	0,625	0,656
(Nuestro) XLMem*	0,767	0,795	0,781	0,707	0,672	0,689	0,649	0,619	0,633
(Nuestro) XLMem*+MTB	0,788	0,709	0,746	0,755	0,616	0,678	0,707	0,584	0,640

Tabla 5

Resultados obtenidos por los diferentes sistemas en la tarea de Extracción de Relaciones. Mejores resultados en cada métrica están marcados en negrita e indican el uso de datos anotados automáticos adicionales para el entrenamiento. El XLMem\*

El sistema es el que se ha enviado a concurso.

podría tener un impacto.

## 5.2. Tarea de extracción de relaciones (B)

En esta parte discutimos los resultados obtenidos por nuestros sistemas de ER durante el desarrollo y pruebas. Comparamos nuestros tres sistemas diferentes con el resto de los principales competidores y evaluamos más exhaustivamente nuestros propios sistemas comparando las curvas de recuperación de precisión y analizando la confusión en la predicción.

La Tabla 5 muestra resultados comparativos (precisión, recuperación y puntuación F1) entre nuestros sistemas y el resto son los principales competidores. Como los organizadores solo informaron el conjunto de prueba, simplemente compare nuestro sistema con el resto en esa partición específica. Los resultados del programa de desarrollo lo siguiente: 1) los datos anotados automáticamente adicionales tienen un efecto positivo en el modelo regularización, y 2) el preentrenamiento de MTB aumenta la precisión al perder la recuperación. El mejor modelo según el conjunto de desarrollo es el XLMem\*, que formó parte del proyecto oficial. Por el contrario, los resultados de las pruebas muestran un comportamiento inesperado. Nuestra hipótesis es que esto se debe a las diferencias en la distribución del tipo de relación de las particiones de desarrollo y prueba (Figura 4). Sin embargo, cada uno de los extractores de relaciones propuestos supera al resto de los sistemas por un gran margen.

La Figura 5 presenta curvas de recuperación de precisión en las categorías de relación de los tres modelos RE. El Las curvas muestran que los sistemas XLMem y XLMem\* funcionan de manera similar a sus sistemas micropromediados. Las curvas están muy cerca, pero no para el XLMem\*+MTB, que la curva se desempeña por debajo del resto. Por otro lado, las curvas sobre categorías de relación muestran que el sistema XLMem\* funciona mejor en las relaciones de roles de acción, y XLMem funciona mejor en las relaciones de dominio general. Las diferencias en el desarrollo y las pruebas también pueden explicarse por la distribución que se muestra en Figura 4. Finalmente, el análisis de los resultados revela que la confusión se localiza en realidad entre la clase negativa (no-relación) y el resto de relaciones positivas (es decir, falsos negativos), y no entre los tipos de relaciones positivas.

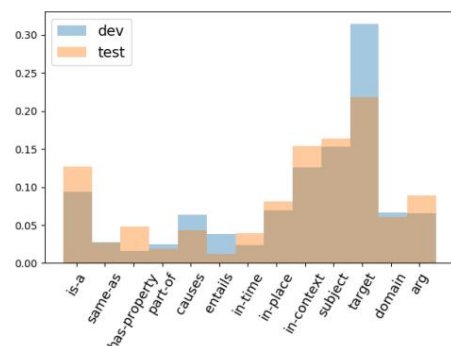


Figura 4: Distribución de relaciones de conjuntos de datos de desarrollo y prueba.

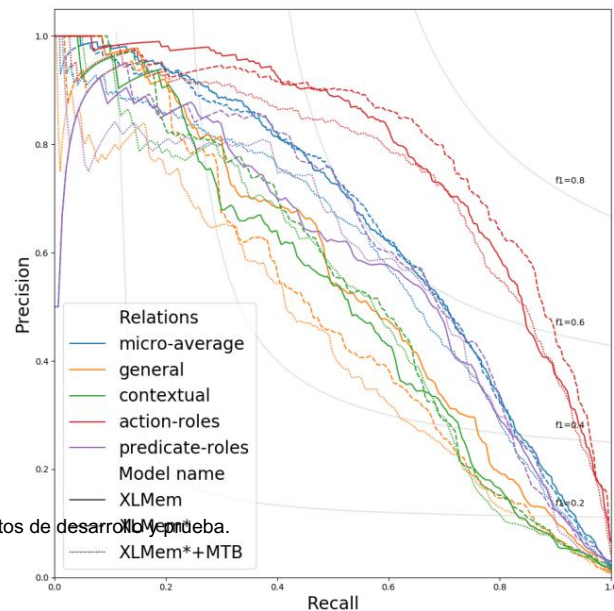


Figura 5: Curvas de precisión/recuperación de los diferentes sistemas.

## 6. Conclusiones

El propósito de este trabajo fue evaluar la viabilidad de diferentes enfoques para el reconocimiento de entidades médicas y la extracción de relaciones para el español. El reconocimiento de entidades se abordó con un etiquetador de secuencia basado en caracteres, y para la extracción de relaciones utilizamos un modelo de lenguaje multilingüe preentrenado de gran tamaño y afinado. El sistema propuesto muestra resultados prometedores. Ocupamos el cuarto lugar en general y obtenemos los mejores resultados para la tarea de extracción de relaciones. En el futuro, planeamos mejorar la parte de reconocimiento de entidades mediante el uso de un LM específico de dominio e investigar más a fondo el uso del método Matching the Blanks como técnica de aumento de datos.

## Referencias

- [1] Piad-Morffis, Y. Gutierrez, H. Cañizares-Díaz, S. Estevez-Velarde, Y. Almeida-Cruz, R. Muñoz y A. Montoyo, Overview of the eHealth Knowledge Discovery Challenge en IberLEF 2020, en : Actas del Foro Ibérico de Evaluación de Lenguas compartido con la 36ª Conferencia de la Sociedad Española para el Procesamiento del Lenguaje Natural, IberLEF@SEPLN 2020, España, septiembre de 2020.
- [2] A. Akbik, D. Blythe, R. Vollgraf, Incorporaciones de cadenas contextuales para etiquetado de secuencias, en: Actas de la 27.ª Conferencia Internacional sobre Lingüística Computacional, Asociación de Lingüística Computacional, Santa Fe, Nuevo México, EE. UU., 2018 , págs. 1638-1649. URL: <https://www.aclweb.org/anthology/C18-1139>.
- [3] G. Lample, A. Conneau, Preentrenamiento del modelo de lenguaje multilingüe, Avances en Neural Sistemas de Procesamiento de Información (NeurIPS) (2019).

- [4] L. Baldini Soares, N. FitzGerald, J. Ling, T. Kwiatkowski, Emparejar los espacios en blanco: similitud distributiva para el aprendizaje de relaciones, en: Actas de la 57ª Reunión Anual de la Asociación de Lingüística Computacional, Asociación de Lingüística Computacional, Florencia, Italia, 2019, págs. 2895–2905. URL: <https://www.aclweb.org/anthology/P19-1279>. doi:10.18653/v1/P19-1279.
- [5] G. Zhou, J. Zhang, J. Su, D. Shen, C. Tan, Reconocimiento de nombres en textos biomédicos: un enfoque de aprendizaje automático, *Bioinformatics* 20 (2004) 1178–1190.
- [6] F. Soares, M. Villegas, A. González-Agirre, M. Krallinger, J. Armengol-Estapé, Incorporaciones de palabras médicas para español: desarrollo y evaluación, en: Actas del 2º Taller clínico de procesamiento del lenguaje natural, Asociación de Lingüística Computacional, Minneapolis, Minnesota, EE. UU., 2019.
- [7] P. Stenetorp, H. Soyer, S. Pyysalo, S. Ananiadou, T. Chikayama, El tamaño (y el dominio) importa: Evaluación de representaciones semánticas del espacio de palabras para texto biomédico, en: Actas del 5º Simposio Internacional sobre Minería Semántica en Biomedicina, Zürich, Suiza, 2012.
- [8] ME Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, Representaciones de palabras contextualizadas profundas, en: Proc. de NAACL, 2018.
- [9] J. Howard, S. Ruder, Ajuste del modelo de lenguaje universal para la clasificación de texto, 2018. arXiv:1801.06146.
- [10] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Entrenamiento previo de transformadores bidireccionales profundos para la comprensión del lenguaje, preimpresión de arXiv arXiv:1810.04805 (2018).
- [11] L. Akhtyamova, P. Martinez, K. Verspoor, J. Cardiff, Prueba de incrustaciones de palabras contextualizadas para mejorar ner en narrativas de casos clínicos en español, preimpresión de BMC Medical Informatics and Decision Making (2020). doi:10.21203/rs.2.22697/v1.
- [12] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, CH So, J. Kang, BioBERT: un modelo de representación de lenguaje biomédico previamente entrenado para la minería de textos biomédicos, *Bioinformatics* (2019). URL: <https://doi.org/10.1093/bioinformatics/btz682>. doi:10.1093/bioinformática/btz682.
- [13] Y. Si, J. Wang, H. Xu, K. Roberts, Mejora de la extracción de conceptos clínicos con incrustaciones contextuales, *Revista de la Asociación Estadounidense de Informática Médica* 26 (2019) 1297–1304. URL: <http://dx.doi.org/10.1093/jamia/ocz096>. doi:10.1093/jamia/ocz096.
- [14] G. Sheikhshabbafghi, I. Birol, A. Sarkar, Las incorporaciones de tokens con reconocimiento del contexto en el dominio mejoran el reconocimiento de entidades biomédicas con nombre, en: Actas del Noveno Taller Internacional sobre Minería de Textos de Salud y Análisis de Información, Asociación de Lingüística Computacional, Bruselas, Bélgica, 2018, págs. 160-164. URL: <https://www.aclweb.org/anthology/W18-5618>. doi:10.18653/v1/W18-5618.
- [15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, AN Gomez, Ł. Kaiser, I. Polosukhin, La atención es todo lo que necesita, en: Avances en los sistemas de procesamiento de información neuronal, 2017, págs. 5998–6008.
- [16] ME Peters, M. Neumann, RL Logan, R. Schwartz, V. Joshi, S. Singh, NA Smith, Representaciones de palabras contextuales mejoradas por conocimiento, en: EMNLP, 2019.
- [17] M. Joshi, D. Chen, Y. Liu, DS Weld, L. Zettlemoyer, O. Levy, SpanBERT: Mejora del entrenamiento previo mediante la representación y predicción de tramos, preimpresión de arXiv arXiv:1907.10529 (2019).
- [18] M. Mintz, S. Bills, R. Snow, D. Jurafsky, Supervisión distante para la extracción de relaciones sin datos etiquetados, en: Actas de la Conferencia Conjunta de la 47ª Reunión Anual de la

- ACL y la Cuarta Conferencia Internacional Conjunta sobre Procesamiento del Lenguaje Natural de la AFNLP, Asociación de Lingüística Computacional, Suntec, Singapur, 2009, págs. URL: <https://www.aclweb.org/anthology/P09-1113>.
- [19] O. Sainz, O. López de Lacalle, I. Aldabe, M. Maritxalar, Supervisión a distancia adaptada al dominio para la extracción de relaciones con motivación pedagógica, en: Actas de la 12ª Conferencia de Evaluación y Recursos Lingüísticos, Asociación Europea de Recursos Lingüísticos, Marsella, Francia, 2020, págs. 2213–2222. URL: <https://www.aclweb.org/anthology/2020.lrec-1.270>.
- [20] R. Hoffmann, C. Zhang, X. Ling, L. Zettlemoyer, DS Weld, Supervisión débil basada en el conocimiento para la extracción de información de relaciones superpuestas, en: Actas de la 49ª Reunión Anual de la Asociación de Lingüística Computacional: Humano Tecnologías del lenguaje , Asociación de Lingüística Computacional, Portland, Oregon, EE. UU., 2011, págs. 541–550. URL: <https://www.aclweb.org/anthology/P11-1055>.
- [21] E. Grave, P. Bojanowski, P. Gupta, A. Joulin, T. Mikolov, Aprendizaje de vectores de palabras para 157 idiomas, en: Actas de la Conferencia Internacional sobre Evaluación y Recursos Lingüísticos (LREC 2018), 2018.
- [22] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Brew, Los transformadores de Huggingface: Procesamiento de lenguaje natural de última generación , ArXiv abs/1910.03771 (2019).
- [23] L. Padró, E. Stanilovsky, Freeling 3.0: Hacia una multilingüidad más amplia, en: Actas de la Conferencia de Evaluación y Recursos Lingüísticos (LREC 2012), ELRA, Estambul, Turquía, 2012.