El olvido catastrófico en el contexto del reconocimiento de entidades nombradas (NER) en Inteligencia Artificial (IA) se refiere a la pérdida de habilidades o conocimientos previamente adquiridos por un modelo de IA cuando se le enseña nuevas tareas, especialmente en el ámbito del procesamiento del lenguaje natural (NLP). Este fenómeno es particularmente relevante en el NER, donde el modelo de IA se entrena para identificar y clasificar entidades como nombres de personas, organizaciones, lugares, entre otros, en textos **REFERENCIA**.

Cuando un modelo de IA se entrena para realizar tareas de NER, puede sufrir el olvido catastrófico si se le enseña a realizar otras tareas de NLP, como la clasificación de texto o la traducción automática. Esto significa que, aunque el modelo pueda aprender nuevas habilidades, puede perder su capacidad para identificar y clasificar correctamente las entidades nombradas en textos, lo que afecta su rendimiento en la tarea original de NER. [1]

El equipo de la Universidad de Catania, Italia, liderado por Concetto Spampinato, ha trabajado para abordar este problema mediante el desarrollo de un método de aprendizaje consolidado de vigilia y sueño (WSCL). Este enfoque intenta replicar el proceso de aprendizaje humano, donde los humanos afianzan conceptos y aprendizajes durante el sueño, lo que les permite recordar cosas mejor. En el contexto de la IA, el WSCL combina una fase de entrenamiento "despierto", donde el modelo se entrena con un conjunto de datos, y una fase de "sueño", donde el modelo analiza los datos de la fase despierta y también resúmenes de cargas de datos anteriores para evitar el olvido catastrófico.

Los resultados de las pruebas realizadas por este equipo mostraron mejoras significativas en la precisión del reconocimiento de entidades nombradas, con incrementos entre un 2 y un 12% en comparación con el método de aprendizaje tradicional.

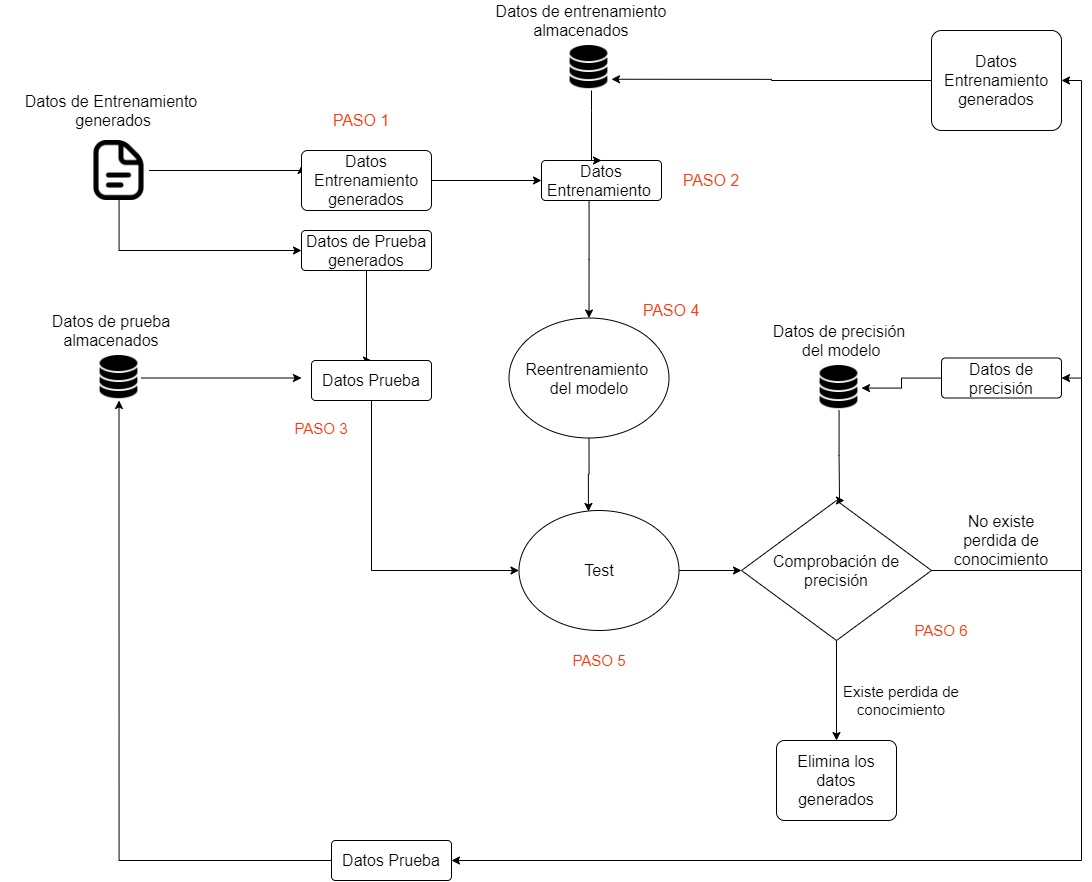
## Algoritmo Propuesto para resolver esta problemática

Para resolver el problema anterior adaptado a la problemática de investigación se diseñó el siguiente algoritmo ver Fig1:

Entrada: X, Y

Salida: Z

1. primero se toman los datos generados por el modelo de llama2 para y se deviden en una proporción de 75% para datos de entrenamiento y 25% para datos de prueba.
2. Los datos de entrenamiento generados se concatenan con los datos de entrenamiento almacenados en la base de datos de Elasticsearch al igual que los datos de prueba con su índice correspondiente.
3. Se reentrenamos el modelo de Spacy con la unión de los datos nuevos y viejos. El resultado le comprobamos la precisión, la exactitud, F1-Score, y el recall.
4. Comparamos estas medidas con las anteriores del modelo almacenadas en la base de datos, si no existe pérdida de conocimiento entonces guardamos los datos de entrenamiento, los de prueba, y las métricas del modelo calculadas.
5. Si se pierde conocimiento se le envía un mensaje de error al usuario para que vuelva a generar los datos.



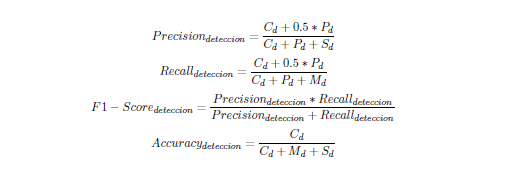
**Fig 1**

## Medidas de desempeño

Para la detección de entidades, se diferenciarán cuatro casos:

* Correcta (Cd): Una detección será correcta cuando sobre en una frase se encuentra una anotación con el mismo inicio y final.
* Parcial (Pd): Una detección es parcial si en una frase se solapan las detecciones. Por ejemplo, que comiencen en la misma posición de inicio, pero finalicen en una posición final diferente. Uno de los casos en los que se puede producir es cuando el modelo nuevo predice una entidad en varias partes, mientras que en el gold standard es solo una.
* Missing (Md): Una detección será clasificada como missing (falta), cuando se encuentra en el gold standard pero no en las detectadas por el nuevo modelo.
* Spurious (Sd): Este último caso sería la inversa del missing. Se da cuando el modelo ha detectado una entidad que no está en el gold standard.

Con estos cuatros elementos, se calcularán la precisión (precision), exhaustividad (recall), puntuación F1 (F1 score) y la exactitud (accuracy) del modelo. Para ello, se utilizan las siguientes fórmulas [2]:



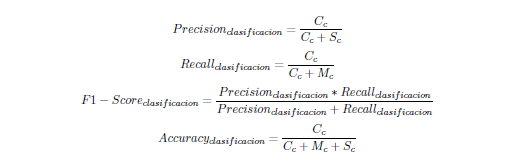
Por otro lado, para la clasificación de entidades, se diferenciarán otros tres casos:

* Correcta (*Cc*): Una clasificación será correcta cuando en una frase se encuentra una categoría del *gold standard* en las anotaciones generadas por el modelo.
* *Missing* (*Mc*): Una clasificación será *missing* (falta), cuando se encuentra en el *gold standard* pero no en las detectadas por el nuevo modelo.
* *Spurious* (*Sc*): Este último caso sería la inversa del *missing*. Se da cuando el modelo ha clasificado una entidad que no está en el *gold standard*.

Con estos tres elementos, se calcularán la precisión (*precision*), exhaustividad (*recall*),

puntuación F1 (*F1 score*) y la exactitud (*accuracy*) del modelo. Para ello, se utilizan las

siguientes fórmulas.



Para el set de datos de comprobación con las nuevas entidades a reconocer se calcularon las siguientes métricas.

|  |  |  |
| --- | --- | --- |
| Métrica | Modelo Original | Modelo Reentrenado |
| Precisión | 0.7833 | 0.882 |
| Recall | 0.5465 | 0.849 |
| F1-Score | 0.321 | 0.432 |
| Accuracy | 0.4791 | 0.762 |

**Bibliografía**

[1] «Qué es el “olvido catastrófico”, perdición de la inteligencia artificial». Accedido: 6 de abril de 2024. [En línea]. Disponible en: https://www.abc.es/ciencia/abci-olvido-catastrofico-perdicion-inteligencia-artificial-201905281851\_noticia.html

[2] A. S. Pérez, «Generación de Datos Sintéticos para el Refuerzo de Modelos de Aprendizaje Automático en Entornos Reales para la Extracción de Información».

**Usa estas referencias:**

Song, Xiang, et al. "Overcoming Catastrophic Forgetting for Multi-Label Class-Incremental Learning." *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2024.

Liu, Xinchen, et al. "Overcoming catastrophic forgetting with classifier expander." *Asian Conference on Machine Learning*. PMLR, 2024.