Reconocimiento de Entidades y Relaciones en Descripciones de Pinturas Usando Procesamiento de Lenguaje Natural

Felipe Cruz *fcruzv@unal.edu.co,* David Casallas, *dcasallasm@unal.edu.co,* Luis Rodríguez, *luiarodriguezper@unal.edu.co*

***Abstract*—Este trabajo presenta un sistema de procesamiento de lenguaje natural para reconocer entidades y relaciones en descripciones de pinturas, utilizando modelos preentrenados en español. Las descripciones se limpiaron, tradujeron y etiquetaron automáticamente, y se utilizó validación cruzada para seleccionar el modelo de etiquetado más preciso. Las entidades fueron extraídas como sustantivos y pronombres, excluyendo términos posicionales, mientras que las relaciones se identificaron mediante patrones específicos de etiquetas gramaticales. Finalmente, los resultados se almacenaron en formato `.json`, estructurando la información en términos de objetos y sus relaciones para facilitar su análisis posterior.**

***Index Terms*—Procesamiento de Lenguaje Natural, Reconocimiento de Entidades, Relaciones Semánticas, Spacy, Validación Cruzada, Etiquetado de Texto, Patrones Gramaticales, Formato JSON.**

I. Introducción

El reconocimiento de entidades y la extracción de relaciones son tareas fundamentales en el campo del procesamiento de lenguaje natural (PLN), utilizadas para estructurar información no estructurada proveniente de textos. Este proyecto se centra en la creación de un modelo capaz de identificar entidades y sus relaciones en descripciones textuales asociadas a un conjunto de imágenes.

El modelo desarrollado se basa en la identificación de patrones lingüísticos mediante el etiquetado de partes del discurso (POS, por sus siglas en inglés) utilizando herramientas como Spacy y bases de datos preexistentes, optimizando su rendimiento para el idioma español. Estas descripciones incluyen referencias a elementos clave presentes en las imágenes, así como sus posiciones y relaciones espaciales, permitiendo construir una representación estructurada y comprensible.

Con este propósito, se diseñaron y aplicaron técnicas de preprocesamiento, traducción y formateo de datos para entrenar un modelo robusto. El sistema resultante busca ser una herramienta versátil que permita no solo identificar entidades, sino también establecer vínculos semánticos entre ellas, abriendo nuevas posibilidades en aplicaciones como la interpretación de imágenes y la generación de conocimiento estructurado.

II. Pre-Procesamiento

La base de datos utilizada, obtenida de Huggingface [1], consiste en un conjunto de imágenes de pinturas acompañadas de descripciones detalladas. Estas descripciones proporcionan información sobre los elementos presentes en cada pintura, así como su posición dentro del espacio representado. Sin embargo, algunas partes del texto, como oraciones iniciales repetitivas y metadatos adicionales, no aportan valor al objetivo del proyecto y fueron consideradas para su eliminación.

A. Filtrado y Limpieza

Para mejorar la calidad de las descripciones y enfocarse en la información relevante para el reconocimiento de entidades y relaciones, se aplicaron técnicas de limpieza textual mediante expresiones regulares. Este proceso se enfocó en eliminar frases y patrones recurrentes que no aportaban información útil al análisis, como referencias generales a la pintura misma o a los artistas responsables de las obras.

Se eliminaron frases iniciales comunes como “The image shows a painting of”, así como fragmentos relacionados con los autores de las obras, por ejemplo, “Painted by”. Además, se corrigieron redundancias y errores de formato, tales como espacios dobles, puntos innecesarios y comas mal colocadas. Los ejemplos del proceso de limpieza se resumen en la Tabla I, donde se presentan los textos originales junto con los resultados después de aplicar estas reglas, así como una breve descripción de cada caso. Este preprocesamiento permitió reducir el ruido en los datos y garantizar que las descripciones resultantes se centraran exclusivamente en los elementos esenciales presentes en cada pintura.

TABLE I
EJEMPLOS DEL PROCESO DE LIMPIEZA DE LAS DESCRIPCIONES

Texto Original	**Texto Limpio**	**Descripción de la Limpieza**
The image shows a painting of a beautiful sunset over the mountains.	A beautiful sunset over the mountains.	Eliminación de introducciones estándar como "The image shows a painting of".
Painted by John Doe. The painting is a representation of a rural village.	A representation of a rural village.	Eliminación de referencias al artista, como "Painted by".
The painting is a stunning depiction of a forest in autumn.	A stunning depiction of a forest in autumn.	Eliminación de frases como "The painting is" al inicio de las descripciones.
The painting of the castle is breathtaking.	A breathtaking castle.	Reducción de redundancias como "The painting of".
Of the painting, the vibrant colors stand out.	The vibrant colors stand out.	Eliminación de expresiones como "Of the painting".
A group of people, painted in vivid colors, walking through the park.	A group of people walking through the park.	Limpieza de redundancias como "painted in vivid colors".
The painting, a representation of a cityscape at dusk, captures the moment perfectly.	A representation of a cityscape at dusk captures the moment perfectly.	Reducción de descripciones innecesarias relacionadas con "The painting".

B. Traducción

Una vez finalizado el proceso de limpieza, se procedió a traducir las descripciones, originalmente en inglés, al español, dado que el objetivo del proyecto era construir un modelo adaptado a este idioma. Para ello, se utilizó la API de Google Translator [2], aplicando la traducción sobre el texto limpio. Sin embargo, debido a las limitaciones de uso de la API, que restringe el número de peticiones permitidas en un corto lapso de tiempo, fue necesario realizar la traducción en lotes de hasta 2.000 textos por operación.

Pese a estos esfuerzos, las restricciones temporales y técnicas llevaron a priorizar un subconjunto de los datos, reduciendo el volumen de las 46.000 descripciones originales a 10.329 textos traducidos. Este conjunto final de datos en español se utilizó para entrenar el modelo.

III. ENTRENAMIENTO

El entrenamiento del modelo de etiquetado de partes del discurso (POS, por sus siglas en inglés) constituyó una de las etapas principales del proyecto. El objetivo fue desarrollar un modelo en español capaz de diferenciar las palabras del texto según su función gramatical, como sustantivos, verbos y adjetivos, entre otras. Esta etapa combinó el uso de modelos preexistentes para etiquetar datos, la aplicación de técnicas de validación cruzada (*Cross Validation*) y un proceso de selección del mejor modelo, asegurando así un rendimiento óptimo en la tarea objetivo.

Para ello, se partió del conjunto de datos previamente procesado y traducido, el cual se preparó en un formato compatible con los requisitos del entrenamiento del modelo de Spacy [3]. En las subsecciones siguientes, se detalla el procedimiento empleado para etiquetar los textos utilizando un modelo preentrenado, los pasos realizados para entrenar y validar los modelos utilizando *K-Fold Cross Validation*, y los criterios para seleccionar el modelo final con mejor desempeño.

A. Etiquetar textos con modelo preexistente

Dado el tamaño considerable del conjunto de datos procesado y traducido, realizar una etiquetación manual de las palabras era inviable. Para abordar este desafío, se decidió utilizar un modelo preexistente de etiquetado de partes del discurso (POS) en español, asumiendo que las etiquetas proporcionadas por este modelo eran correctas.

El modelo seleccionado fue **`es_core_news_md`** [4], una pipeline de Spacy diseñada específicamente para el idioma español. Este modelo ha sido entrenado con un corpus de noticias y es capaz de identificar con alta precisión las categorías gramaticales de las palabras, como sustantivos, verbos y adjetivos, entre otras. Gracias a esta herramienta, se pudieron etiquetar automáticamente los textos del conjunto de datos, lo que permitió preparar un corpus etiquetado listo para las etapas posteriores de entrenamiento.

B. K-Fold Cross Validation para entrenar y validar 10 modelos

Para garantizar un entrenamiento robusto y minimizar el sesgo en la selección del mejor modelo, los 10.329 textos previamente etiquetados fueron divididos en 10 subconjuntos, siguiendo el esquema de *K-Fold Cross Validation* de [5]. En este método, cada subconjunto se utiliza una vez como conjunto de validación, mientras los restantes forman el conjunto de entrenamiento. Esto permitió entrenar y validar 10 modelos diferentes, asegurando que cada modelo fuera evaluado con datos distintos de aquellos usados en su entrenamiento.

El procedimiento se realizó de la siguiente manera: el *modelo 1* se entrenó utilizando los primeros 9 subconjuntos y se validó con el décimo; el *modelo 2* utilizó como validación el segundo subconjunto, mientras los restantes sirvieron para el entrenamiento, y así sucesivamente. Este enfoque permitió maximizar el uso del conjunto de datos y obtener una evaluación completa del rendimiento del modelo en diferentes particiones.

Cada modelo fue entrenado durante 10 épocas (*epochs*) para asegurar una convergencia adecuada. Los detalles sobre el proceso de validación y los criterios de selección del mejor modelo se presentan en la siguiente subsección.

C. Selección del mejor modelo

El método de validación empleado consistió en calcular, para cada modelo y su correspondiente conjunto de datos de validación, el porcentaje de etiquetas correctas por texto. Este porcentaje se obtuvo dividiendo el número de etiquetas correctamente asignadas entre el total de etiquetas presentes en cada texto. Posteriormente, se calculó el promedio de estos valores a lo largo de todos los textos del conjunto de validación, resultando en una métrica que varía entre 0 y 1, donde 1 representa un modelo perfectamente preciso.

Con base en este enfoque, se evaluaron los 10 modelos entrenados, obteniéndose los resultados que se presentan en la Tabla II.

TABLE II
RESULTADOS DE VALIDACIÓN PARA LOS 10 MODELOS

Modelo	Precisión Promedio
Modelo 1	0.998397
Modelo 2	0.997909
Modelo 3	0.998329
Modelo 4	0.997815
Modelo 5	0.997983
Modelo 6	0.998036
Modelo 7	0.998215
Modelo 8	0.998160
Modelo 9	0.998214
Modelo 10	0.998145

A partir de los resultados obtenidos, se seleccionó el modelo con la mayor precisión promedio, el cual fue el *Modelo 1* con una precisión de 0.998397.

IV. DETECCIÓN DE ENTIDADES Y RELACIONES

Una vez seleccionado el mejor modelo de etiquetado de partes del discurso (POS), se procedió a implementar el sistema para la detección de entidades y sus relaciones en los textos. Este proceso incluyó varias etapas: etiquetar el texto utilizando el modelo seleccionado, extraer las entidades identificadas, determinar las relaciones entre ellas y almacenar la información de manera estructurada. Estas etapas garantizan que los textos procesados sean convertidos en una representación comprensible y útil para el análisis.

En las subsecciones siguientes, se detalla cada paso de este proceso, desde la aplicación del modelo hasta el almacenamiento de las entidades y sus relaciones.

A. Etiquetado con el mejor modelo

En esta etapa, se procesaron todos los textos traducidos disponibles, un total cercano a 11.000, dado que no se excluyeron aquellos que no pasaban los filtros previos. Para cada texto, se aplicó el modelo seleccionado como el mejor durante la etapa de entrenamiento. Este modelo realizó el etiquetado de las palabras, asignando categorías gramaticales como sustantivos, verbos y adjetivos, entre otras, de acuerdo con su función dentro del texto.

El resultado de esta fase fue un conjunto de textos etiquetados que serviría como base para las etapas posteriores de extracción de entidades y relaciones.

B. Extraer entidades del texto

Para la extracción de entidades, se recorrieron todas las palabras etiquetadas de los textos procesados, identificando como entidades aquellas que fueron clasificadas como sustantivos (*noun*) o pronombres (*pronoun*). Este enfoque permitió centrarse en elementos textuales relevantes para la representación de los objetos y sujetos mencionados en las descripciones.

Sin embargo, se excluyeron ciertos sustantivos específicos que corresponden a términos que indican relaciones espaciales o posicionales entre objetos, ya que estos son más apropiados para definir relaciones que entidades. Los sustantivos excluidos fueron los siguientes:

- medio
- arriba
- abajo
- izquierda
- derecha
- debajo
- lado
- frente
- detrás
- atrás
- parte
- fondo
- plano

Al filtrar estos sustantivos especiales, se garantizó que las entidades extraídas representaran únicamente objetos y sujetos, mientras que las relaciones entre ellos se abordarían en la siguiente etapa del proceso.

C. Encontrar relaciones entre entidades

Para identificar relaciones entre entidades, se iteró sobre cada par de entidades consecutivas dentro de los textos. Esto significa que las entidades consideradas estaban separadas únicamente por palabras no etiquetadas como entidades; en otras palabras, no podía haber otra entidad entre ellas. Este enfoque permitió identificar relaciones únicamente entre entidades que estaban lógicamente conectadas dentro del texto.

Entre las palabras que separaban dos entidades consecutivas, se buscaron patrones específicos de etiquetas gramaticales generados por el modelo POS. Estos patrones utilizados fueron:

- `ADP`: Preposición o postposición (e.g., ”en”, ”de”).
- `AUX`: Verbo auxiliar (e.g., ”es”, ”está”).
- `ADJ`: Adjetivo (e.g., ”visible”, ”pequeño”).
- `DET`: Determinante (e.g., ”el”, ”la”).
- `NOUN`: Sustantivo, específicamente uno de los sustantivos especiales definidos anteriormente, ya que estos no se consideran entidades sino elementos posicionales/relaciones.

Los patrones evaluados fueron los siguientes:

- `ADP, NOUN, ADP`
- `AUX, ADJ, ADP`
- `ADP, DET`
- `ADP`

Cuando se encontraba un patrón en el texto entre dos entidades, la relación se definía basándose en las palabras correspondientes a dichas etiquetas.

En la Tabla III se presentan ejemplos específicos de los patrones y las relaciones detectadas:

TABLE III
EJEMPLOS DE PATRONES DE ETIQUETAS Y LAS RELACIONES CORRESPONDIENTES

Patrón de etiquetas	Relación detectada
`ADP, NOUN, ADP`	en medio de
`AUX, ADJ, ADP`	es visible en
`ADP, DET`	en la
`ADP`	en

Este enfoque permitió establecer relaciones precisas y relevantes entre las entidades identificadas, asegurando que los elementos posicionales y descriptivos fueran correctamente integrados como relaciones y no confundidos con entidades.

D. Guardar información estructurada

Una vez identificadas las entidades y sus relaciones, la información extraída de cada texto fue almacenada en un archivo en formato `.json`. Este formato estructurado permite representar tanto los objetos detectados como las relaciones entre ellos de manera clara y accesible para futuros análisis o integraciones. La estructura de los archivos generados es la siguiente:

```
{
    "objects": [
        { "id": int, "type": str }
    ],
    "relations": [
        { "type": str, "entity1": int, "
            entity2": int }
    ]
}
```

En esta representación:

- El campo `objects` contiene una lista de entidades, donde cada entidad se define mediante un identificador único (`id`) y su tipo (`type`), que corresponde al texto de la entidad (por ejemplo, `casa`).
- El campo `relations` incluye las relaciones detectadas entre entidades, definidas por su tipo (`type`), que corresponde al texto de la relación (por ejemplo, `en medio de`) y los identificadores (`entity1` y `entity2`) de las entidades conectadas por dicha relación.

El identificador `id` de cada entidad corresponde a un número asignado de manera ascendente a cada palabra del texto, comenzando desde 1. Por su parte, el campo `type` en las entidades y relaciones refleja el texto correspondiente, tal como se identificó en las etapas previas del procesamiento.

Este esquema de almacenamiento garantiza que los datos sean fácilmente reutilizables y adecuados para tareas de análisis posterior, como visualización, evaluación o entrenamiento de modelos adicionales.

V. CONCLUSIONES Y TRABAJO FUTURO

El desarrollo de este proyecto permitió avanzar significativamente en la tarea de reconocimiento de entidades y relaciones en descripciones textuales de imágenes, logrando implementar un sistema basado en procesamiento de lenguaje natural que combina técnicas avanzadas de etiquetado, extracción y análisis semántico. Los resultados obtenidos, con un modelo entrenado que alcanzó una precisión promedio superior al 98%, demuestran la viabilidad del enfoque propuesto para procesar textos y estructurar información de manera eficiente.

Además, se logró identificar entidades relevantes y relaciones significativas, empleando un conjunto de patrones de etiquetas diseñados para capturar conexiones semánticas en los textos. Este trabajo establece una base sólida para futuras investigaciones y desarrollos en el campo del procesamiento de lenguaje natural aplicado a la interpretación de datos textuales y visuales.

A. Trabajo futuro

A pesar de los avances logrados, el proyecto abre múltiples oportunidades para mejorar y expandir las capacidades del sistema desarrollado. Entre las áreas de trabajo futuro se incluyen:

- Mejorar el proceso de limpieza de los textos para garantizar que TODAS las descripciones se enfoquen exclusivamente en el contenido de las pinturas y no incluyan referencias irrelevantes sobre la pintura en sí, su creación o el autor.
- Completar la traducción del dataset completo para aprovechar el total de descripciones disponibles, aumentando así el volumen de datos utilizados en las etapas de entrenamiento y validación.
- Entrenar un modelo con un conjunto de datos más amplio para mejorar su desempeño y robustez frente a variaciones en los textos.
- Validar manualmente las etiquetas asignadas por el modelo preentrenado para confirmar que coincidan con la interpretación de una persona, mejorando la confiabilidad del sistema.
- Ampliar la detección de relaciones para incluir aquellas que involucran más de dos entidades, como en el caso de expresiones que indican que una entidad está ubicada en medio de dos o más entidades.
- Diseñar y agregar más patrones de etiquetas que permitan identificar relaciones adicionales o mejorar la precisión de las relaciones ya detectadas.
- Revisar y extender la lista de sustantivos especiales, incorporando nuevos términos que puedan ser considerados como indicadores de relaciones en lugar de entidades, para mejorar la calidad de las extracciones.

Estas líneas de trabajo futuro permitirán refinar las capacidades del sistema y abrir nuevas posibilidades para su aplicación en contextos más amplios.

REFERENCES

[1] alfredplpl, "ArtBench-PD 256x256 Dataset," [Online]. Available: https://huggingface.co/datasets/alfredplpl/artbench-pd-256x256. [Revisado: Dic. 7, 2024].

[2] SuHun Han, "googletrans: Free and Unlimited Python Library that Implements Google Translate API," [Online]. Available: https://pypi.org/project/googletrans/. [Revisado: Dic. 9, 2024].

[3] Explosion AI, "spaCy: Industrial-Strength Natural Language Processing in Python," [Online]. Available: https://spacy.io. [Revisado: Dic. 9, 2024].

[4] Explosion AI, "spaCy Spanish Model: es_core_news_md," [Online]. Available: https://huggingface.co/spacy/es_core_news_md. [Revisado: Dic. 9, 2024].

[5] scikit-learn developers, "Cross-validation: evaluating estimator performance," [Online]. Available: https://scikit-learn.org/1.5/modules/cross_validation.html. [Revisado: Dic. 9, 2024].

[6] E. León Guzmán, "Preprocesamiento de texto," Presentación de clase, Curso de Procesamiento de Lenguaje Natural, Universidad Nacional de Colombia, Bogotá, Colombia.

[7] E. León Guzmán, "Espacio de representación de texto," Presentación de clase, Curso de Procesamiento de Lenguaje Natural, Universidad Nacional de Colombia, Bogotá, Colombia.

[8] E. León Guzmán, "Aprendizaje Supervisado: Clasificación de textos," Presentación de clase, Curso de Procesamiento de Lenguaje Natural, Universidad Nacional de Colombia, Bogotá, Colombia.

[9] E. León Guzmán, "Part of Speech: Etiquetamiento Gramatical," Presentación de clase, Curso de Procesamiento de Lenguaje Natural, Universidad Nacional de Colombia, Bogotá, Colombia.