

Informe: Propuesta de Solución para Análisis de Cobertura y Pertinencia en documentación de proyectos

Descripción del Problema

En el análisis de textos extensos, como los informes de antecedentes o estudios académicos, identificar y cuantificar variables clave como "cobertura" y "pertinencia" puede ser una tarea ardua y que requiere mucho tiempo. Actualmente, se requiere extraer información muy puntual de proyectos descritos en archivos, se debe leer grandes volúmenes de información para localizar las partes más relevantes, lo que dificulta la eficacia en la cuantificación de datos y extracción de material aprovechable. La falta de herramientas automatizadas que permitan interpretar estos conceptos de manera rápida y precisa genera una barrera para el análisis ágil de la información.

Introducción

Este informe detalla la solución propuesta para automatizar el análisis de variables clave como "cobertura" y "pertinencia" en textos extensos, utilizando técnicas de procesamiento de lenguaje natural (NLP) con la biblioteca spaCy en Python. Esta solución tiene como objetivo agilizar el proceso de interpretación de textos, permitiendo una identificación rápida y precisa de la información aprovechable, sin necesidad de una revisión manual intensiva. A través de la automatización, se busca mejorar la eficiencia en la interpretación de datos y facilitar la cuantificación de dichas variables.

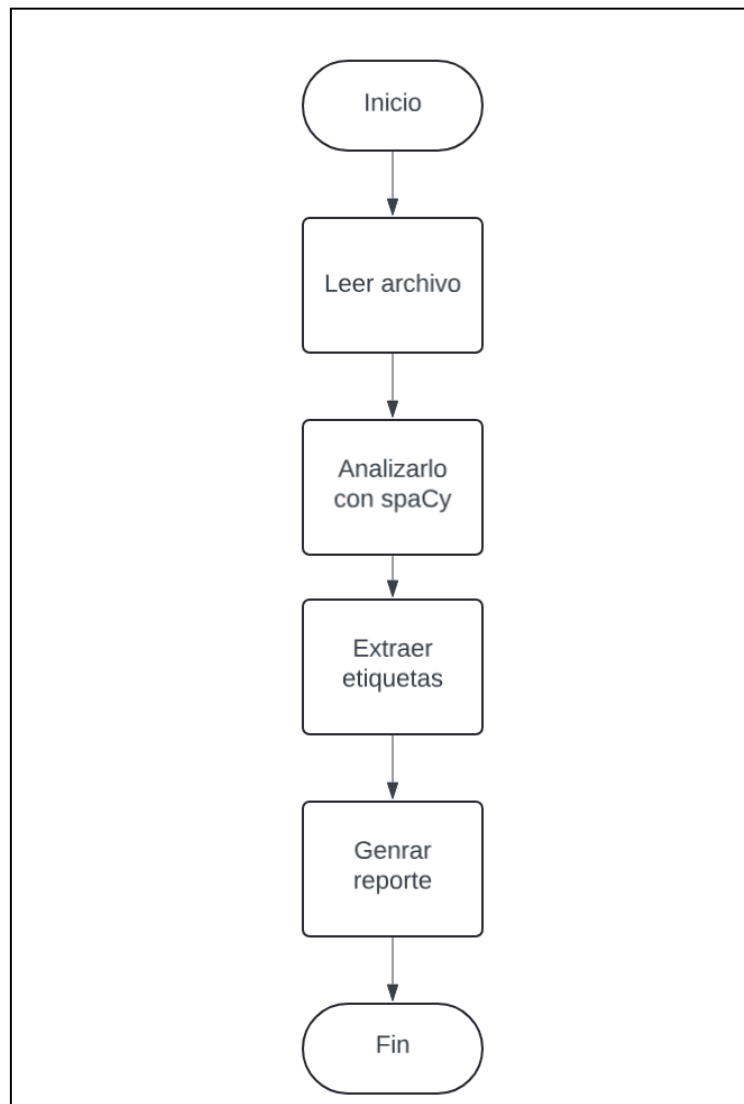
Adicionalmente, se propone el uso de otra biblioteca, como GPT-3 o Hugging Face Transformers, para generar informes automáticos basados en las etiquetas extraídas por spaCy. Esta segunda herramienta interpretará las etiquetas de "cobertura" y "pertinencia" generadas, creando texto en lenguaje natural que presente la información clave de manera estructurada y comprensible, facilitando así la generación automática de reportes detallados.

Descripción de la Solución

La propuesta se basa en el uso de la biblioteca spaCy para el procesamiento de textos en español y la identificación de las entidades clave "cobertura" y "pertinencia". Además, se propone la integración de una segunda herramienta que permita generar informes automáticos a partir de las etiquetas obtenidas en el análisis de texto.

- 1. Extracción de Entidades Clave:** Con spaCy, se extraen las entidades específicas dentro de un texto. Esto implica personalizar el modelo de spaCy para identificar y etiquetar las menciones de "cobertura" y "pertinencia" en documentos extensos. Una vez entrenado el modelo con ejemplos supervisados, el sistema podrá analizar automáticamente nuevos textos y etiquetar las secciones relevantes.
- 2. Entrenamiento del Modelo:** El modelo de spaCy será entrenado con ejemplos etiquetados, lo que le permitirá reconocer de manera precisa las entidades relacionadas con las variables clave. El entrenamiento supervisado asegura que el modelo se ajuste específicamente a las necesidades del análisis de cobertura y pertinencia en el contexto del proyecto.

-
- 3. Automatización del Análisis y Etiquetado:** Una vez entrenado, el modelo de spaCy podrá procesar cualquier texto de entrada, etiquetando automáticamente las secciones correspondientes a "cobertura" y "pertinencia". El resultado será un texto marcado con las etiquetas correspondientes, lo que facilita la interpretación y cuantificación de estas variables sin la necesidad de un análisis manual intensivo.
- 4. Generación de Informes Automáticos:** Tras la obtención de las etiquetas con spaCy, se propone el uso de la biblioteca GPT-3 o similares (por ejemplo, Transformers de Hugging Face) para generar informes desarrollados automáticamente. Estas herramientas permiten interpretar las etiquetas y crear texto en lenguaje natural, formulando reportes comprensibles y detallados basados en los hallazgos del análisis.
- a. **Proceso:** A partir de las etiquetas generadas por spaCy, la segunda librería (como GPT-3) interpretará el contenido etiquetado y generará informes estructurados, proporcionando información más desarrollada como:
 - i. **Ejemplo de Salida:** "En este proyecto, los antecedentes identificados son: [lista de antecedentes]. Los beneficiados por la cobertura del proyecto incluyen: [lista de beneficiados]. La pertinencia del proyecto se refleja en [detalles de pertinencia]."
- 5. Automatización Completa del Proceso:** Al combinar spaCy para la extracción de información y una librería de generación de texto como GPT-3, se puede automatizar el flujo completo de análisis y reporte. Esto permite que el sistema no solo identifique las secciones relevantes del texto, sino que también genere informes narrativos y detallados con la información extraída.

Primer diagrama para la propuesta a solución

Referencias

Explosion AI. "spaCy: Industrial-Strength Natural Language Processing in Python."

Disponible en: <https://spacy.io>.

OpenAI. "GPT-3: Language Models are Few-Shot Learners." Disponible en:

<https://openai.com/research/gpt-3>.

Hugging Face. "Transformers: State-of-the-art Natural Language Processing for

PyTorch, TensorFlow, and JAX." Disponible en: <https://huggingface.co/transformers>.