

Propuesta de proyecto de Fin de Carrera

Detección automática de sellos

Resumen

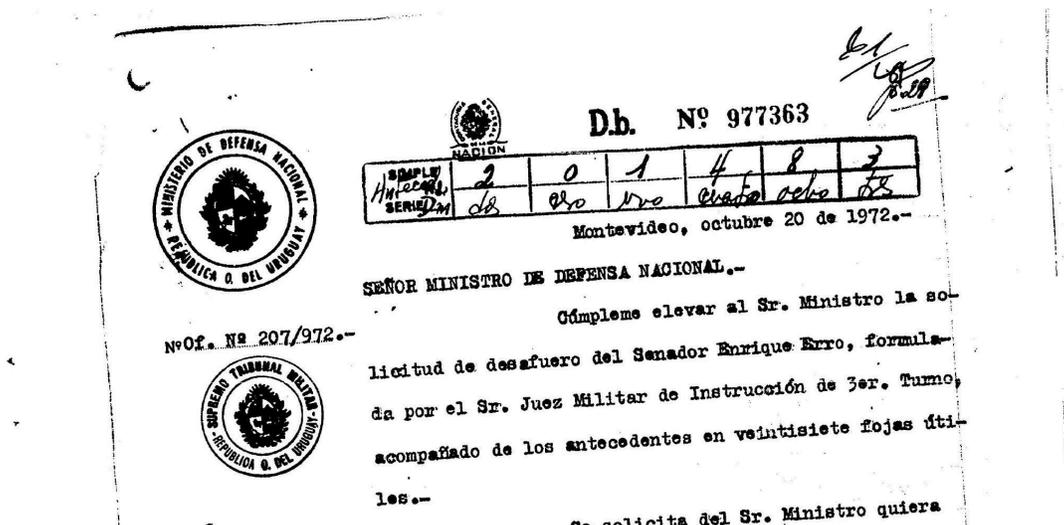
El presente proyecto trata del estudio, implementación y evaluación sistemática de métodos para la detección y extracción de sellos en documentos escaneados. Se estudiarán los métodos existentes en la literatura, se implementará aquellos relevantes que no tengan implementación disponible, y se evaluarán sobre dos conjuntos de datos: a) conjuntos de evaluación de uso académico, público, como [PubLayNet](#) y b) los documentos del Proyecto Cruzar.uy, que incluyen el Archivo Berruti y otros archivos digitalizados de organismos represivos que operaron durante la dictadura militar en Uruguay (1973-1984).

Este proyecto es la continuación de una primera aproximación al problema realizada en el año 2024. Se podrá utilizar todos los resultados y la documentación de dicho proyecto como punto de partida.

El objetivo en esta segunda etapa es perfeccionar la interfaz de modo que quede utilizable para el proyecto Cruzar, e incorporar otros métodos no explorados en el proyecto original.

Antecedentes

La detección de sellos (en inglés *stamps* o *seals*) es un problema muy relevante en el abanico de problemas relacionados con el análisis automático de documentos digitalizados. Abajo vemos un ejemplo del archivo Berruti



El problema planteado tiene especial relevancia en el contexto del proyecto Cruzar.uy, que consiste en varios millones de documentos que deben ser procesados de manera automática, rápida y eficaz. La detección de sellos es muy importante por dos razones: a) por la propia identificación de dichos sellos como información importante de los documentos y b) para quitarlos del documento previo a su transcripción, ya que su presencia confunde y dificulta la transcripción automática.

Resultados esperados

La finalización del proyecto implica lograr dos objetivos bien concretos:

1. Una evaluación sistemática de los métodos más relevantes de la literatura
2. La implementación correcta, eficiente y con interfaz de línea de comandos de un método de detección y extracción automática de sellos en documentos escaneados.

Interesados dirigirse a Ignacio Ramírez, nacho@fing.edu.uy