

Propuesta de objetivo para proyecto de colaboración con el Instituto Milenio Fundamento de los Datos:

La búsqueda de víctimas de desaparición forzada, en lo que dice relación con archivos y bases de datos, se enfrenta a tres desafíos, todos interdependientes entre si.

1. El primero de ellos, consiste en diseñar una base de datos unificada, que contenga todos los campos de información necesaria para la búsqueda, y que permita generar la mayor cantidad de cruces de información posible sobre grandes cantidades de datos, y de la mejor calidad posible.
2. El segundo de ellos, consiste en poblar esa base de datos i) con la información ya sistematizada durante las décadas anteriores (información que no se puede perder); y ii) con nueva información, asegurando la validez de los registros y el respaldo documental del dato. Este último aspecto presenta la particular dificultad de que las cantidades de archivos existentes susceptibles de ser analizados, revisados, descritos e incorporados a una base de datos, excede las capacidades de cualquier grupo humano que se disponga acometer dicha tarea manualmente.
3. El tercer desafío consiste en contar con un software capaz de procesar toda la información sistematizada en la base de datos, y generar la visualización de todos los cruces posibles de información de la forma más útil para la búsqueda, y que sea interoperable entre todos los actores involucrados en la búsqueda.

Al tenor de las conversaciones sostenidas con el Instituto Milenio Fundamento de los Datos, el principal aporte que podemos identificar para el PNB se encuentra en desarrollar una herramienta para agilizar los procesos descritos en el ii) del 2. antes señalados. Esto es confirmado por la experiencia del Plan Nacional de Búsqueda Mexicano, y los aprendizajes de la búsqueda de los casos de desaparición forzada en la contrainsurgencia de dicho país utilizando machine learning, procesamiento de lenguaje natural y ciencia de datos.

En resumen, un objetivo que le interesa a la Subsecretaría de DDHH para el PNB es desarrollar una herramienta, que a través de OCR y procesamiento de lenguaje natural, logre identificar documentos que contengan información relevante para la búsqueda, que un analista luego pueda revisar y validar su interés, cargando su descripción documental en el repositorio de la base de datos, e incorporando la información útil en los campos correspondientes de la base de datos. De esta forma, sería posible agilizar el procesamiento de las grandes cantidades de archivos potencialmente útiles, de forma de hacer un uso más eficiente de los recursos humanos, centrándolos en analizar documentos específicos con potencial utilidad, y no toda la documentación en estrategia de rastrillo.