Personalmente me sentí feliz cuando fui convocada por el Consorcio Internacional de Periodistas de Investigación, para integrar el equipo de Panamá Papers y Paradise Papers.

En este proyecto, los desarrolladores usaron el concepto de Machine Learning para analizar millones de documentos a través de un motor de búsqueda personalizado.
Según cuenta Rigoberto Carvajal en este artículo: https://www.icij.org/blog/2018/08/how-machine-learning-is-revolutionizing-journalism/ en filtraciones como Paradise Papers, el desafío fue tratar con millones de documentos (incluidos archivos PDF, fotos y correos electrónicos) que las plataformas tradicionales como Excel no pueden procesar.

Esto se conoce como Big Data, donde los grandes volúmenes de datos a menudo no organizados deben ser organizados en conjuntos estructurados.

El equipo de ICIJ desarrolló una herramienta de reconocimiento óptico de caracteres, denominada Extract, capaz de reconocer e indexar el texto utilizando la potencia de hasta 30 servidores o más.

También se usó Talend Studio. Es un software de transformación y análisis de datos basado en componentes visuales que se conecta para crear un flujo.

"Primero, automatizamos una búsqueda y almacenamos los resultados creando un trabajo Talend que busca un tema, nombre de organización o individuo que estamos investigando, como 'Glencore'. La búsqueda arrojará resultados y otra información del texto de los documentos. y sus metadatos, como la identificación del documento, el nombre del archivo, la raíz del archivo, la extensión, el tamaño en bytes", explica Rigoberto en el artículo.

Los resultados de búsqueda se almacenan en Solr, una base de datos que reconoce la relación entre los elementos.

El siguiente paso fue agrupar los documentos. La agrupación es una técnica que nos permite agrupar cosas similares. En lugar de desplazarse por una gran cantidad de archivos PDF, es útil crear grupos de documentos por tema o por tipo de documento, para que el periodista pueda acceder a documentos similares de una sola vez, como todos los archivos PDF relacionados con las transferencias de fondos.

"Utilizamos RapidMiner para procesar el texto y los metadatos de documentos y crear clusters basados en palabras y frases comunes. RapidMiner es una poderosa herramienta que facilita la implementación de algoritmos de minería de datos, y también utiliza un espacio de trabajo visual en lugar de un editor de texto monótono"

**¿Qué tan efectiva es la IA en las redacciones?**

Para que un sistema de Inteligencia Artificial pueda ser considerado efectivo, tiene que pasar el denominado test de Turing. Esta prueba, creada por Alan Turing, comprueba la habilidad de una máquina de exhibir un comportamiento similar o igual al de un humano. Esta prueba resulta de una importancia vital cuando hablamos de algoritmos creativos, ya que los consumidores no quieren contenido que ha sido creado por un bot, ya que tradicionalmente se considera que no pueden conectar con el nivel emocional de los humanos.

La mayoría de herramientas de Inteligencia Artificial no pueden escribir al mismo nivel que un humano

Actualmente la Inteligencia Artificial se utiliza de forma limitada en las redacciones, pero es de esperar que en el futuro cercano este uso aumente y se generen nuevas formas de colaboración entre periodistas y robots.

Casos:

The New York Times está usando el aprendizaje automático (machine learning) para buscar patrones en los datos de financiación de campañas, para optimizar sus resultados.

Los Angeles Times construyó el llamado "Bot Quake" para enviar, sin intervención humana, actualizaciones en el momento en el que se detecta un terremoto en la ciudad y sus alrededores.

Associated Press lleva tres años utilizando *Automated Insights* para generar presentaciones de cualquier tipo: desde informes de ganancias de empresas públicas hasta clasificaciones de las ligas menores de béisbol.

https://automatedinsights.com/

Automated Insights es una compañía de tecnología con sede en los Estados Unidos que se especializa en software de generación de lenguaje natural que convierte el big data en narrativas legibles.

Los directivos de la agencia sostienen que la automatización de este trabajo le ha permitido a la plantilla del medio contar con un 20% más de tiempo, que está siendo invertido en la elaboración de reportajes más extensos y profundos.