

LA NACION

Caso LNData con Inteligencia Artificial.

Resumen:

El diario La Nación de Argentina, fundado hace casi 150 años por el expresidente Bartolomé Mitre, está entre los periódicos más prestigiosos y antiguos de habla hispana.

En diciembre de 1995 se lanzó *lanacion.com.ar*. Desde 2011 cuenta en su redacción con un área específica, **LNData**, para realizar Periodismo de Datos. LNData ha recibido una gran cantidad de premios internacionales tanto por innovación tecnológica, como por metodologías aplicadas y por resonantes hallazgos periodísticos, y es considerado como uno de los referentes mundiales en la disciplina.

Desde sus primeros pasos, LNData desarrolló novedosos métodos de recolección automática y conversión de documentos desde centenares de fuentes oficiales, complementada con procesos manuales de adaptación y transformación de los mismos, para ponerlos en una plataforma a disposición de los periodistas que utilizan los varios millones de documentos contenidos en dichos data sets para investigar y lograr un *storytelling*.

LNData participó activamente de investigaciones internacionales encaradas por el Consorcio Internacional de Periodismo de Investigación (ICIJ) con sede en Washington, como SwissLeaks, Panamá Papers y Paradise Papers.

Actualmente LNData se encuentra abocado en un proceso de creación de una herramienta basada en .Net/SQL que concentra todo el ciclo de vida de los datos, e incorpora técnicas de IA y Machine Learning para alguna de sus etapas.

En Julio de 2018 comenzó a trabajar con el equipo de CSE de Microsoft para desarrollar una solución que utiliza tres tecnologías: NLP, Machine Learning y Graph Data Base, para analizar y establecer relaciones entre datos (o entidades) y su visualización, facilitando las tareas investigativas de los periodistas.

Esta implementación permitirá a La Nación mantenerse a la vanguardia del periodismo de datos y de investigación al permitir establecer vínculos entre millones de entidades para construir nuevas historias y develar información que tenga real impacto en la ciudadanía, en la política y en mejorar la sociedad.

¿Cómo trabajan? Desde el comienzo, obtienen y actualizan diariamente información proveniente de diferentes fuentes de acceso público (o hechas públicas, como Wikileaks, LeakyMails, mails de Jaime, etc.), y luego de adaptarlas (transformación de textos escaneados en textos reales, Tablas o cuadros en planillas Excel, etc.), para volcarlos luego en una base de datos para, con asistencia de integrantes de LNData que conocieran tanto las bases disponibles, como la forma de consultarlas, realizar las investigaciones.

Este trabajo se realizaba de una forma relativamente artesanal, y en el último año encararon la máxima automatización posible, y además notificar a diferentes periodistas (mediante LNDATA ALERTS) frente a la llegada de un documento (con el link a la fuente original) que podría ser de su interés.

El desafío que tiene LN Data es poder obtener datos de millones de documentos y conectarlos, para generar una estructura de datos que permita establecer relaciones entre entidades y diferentes sucesos. Esto permitirá que los periodistas no tengan que usar su tiempo para buscar vínculos.

Solución:

Desde marzo de 2018, el equipo de CSE de Microsoft Argentina comenzó a trabajar con el área de LNData del diario en un proyecto que consistió en desarrollar una herramienta innovadora capaz de procesar y establecer relaciones entre millones de datos de los diferentes datasets con los que trabajan.

Esta implementación va a permitirle a La Nación mantenerse a la vanguardia del periodismo de datos al poder acceder y procesar millones de datos que le permitan establecer vínculos para construir nuevas historias y develar información que tenga real impacto en la ciudadanía, en la política y en mejorar la sociedad.

¿Cómo funciona?

Es una solución inédita porque combina tres tecnologías, **Natural Language Processing** (identificar y reconocer entidades nombradas), **Machine Learning** (capacitan a la máquina para el NLP, por ejemplo, que pueda identificar que es un “empresa”, un “domicilio”, “una personas”, etc. dentro de un boletín oficial) y **Graph Data Base** (establece modelos de grafos que permite encontrar y hacer evidente las relaciones entre entidades nombradas dentro de millones de documentos de más de 250 fuentes).

La solución cuenta con un motor SQL server soportado por la nube de Microsoft para reconocer entidades que identifican cuatro tipos diferentes (personas, empresas y domicilios) y marca relaciones entre ellos. Para esto usa nodos y gráficos de SQL. Y es escalable por diseño.

Otra facilidad de esta solución es un novedoso sistema para la administración de las alertas (LNDATA ALERTS) que permitirá a cada periodista definir, incluso desde su celular, los criterios por fuente y por contenido para recibir las.

Conclusión:

Esta nueva herramienta permitirá potenciar la tarea de los periodistas, brindarles la posibilidad de mantenerse informados proactivamente sobre temas de su interés cuando aparece alguna información relacionada con casos periodísticos o judiciales con los que estuvo o está trabajando.