



LABORATORIO NACIONAL DE INFORMÁTICA AVANZADA, A.C.

Détox de código: Limpieza de datos sin dolor con Open Refine



Inyección
educativa



Détox de código: Limpieza de datos sin dolor con Open Refine



Público Objetivo: Personas sin experiencia previa en análisis de datos que:

- Trabajan con datos desorganizados (en Excel, CSV, etc.).
- Quieren aprender a estandarizar, limpiar y transformar datos sin programar.
- Buscan automatizar procesos repetitivos de limpieza.

Descripción: Aprende a limpiar y transformar datos masivos con OpenRefine (herramienta gratuita). Elimina duplicados, corrige errores y normaliza datos para análisis confiables sin código.



Objetivo del Curso:

- Diagnosticar problemas comunes en datos sucios (errores, duplicados, formatos inconsistentes).
- Usar OpenRefine desde cero (instalación, interfaz y flujo de trabajo) para limpiar y transformar datos.
- Técnicas de normalización de textos, fechas y categorías
- Exportar datos limpios en formatos útiles (Excel, CSV) rores en futuros proyectos.



Prerrequisitos:

- Tener conexión a internet.

Temas por sesión:



Sesión 1:

Introducción a la Limpieza de Datos

¿Por qué limpiar datos? (Ejemplos de errores comunes).

Instalación y interfaz de OpenRefine.

Importar datos desde Excel, CSV o URLs.



Sesión 2:

Transformaciones Básicas

Normalización de textos (quitar espacios, mayúsculas, caracteres raros).

Dividir columnas y fusionar datos.

Uso de expresiones GREL (lenguaje simple de OpenRefine).



Détox de código: Limpieza de datos sin dolor con Open Refine

Temas por sesión:

Sesión 3:



Limpieza Avanzada

Eliminar duplicados inteligentemente.

Clustering automático para agrupar valores similares (ej: "Café", "Cafe", "CAFÉ").

Reconocimiento de patrones con expresiones regulares (regex básico).

Sesión 4:



Exportar y Automatizar

Exportar datos limpios (Excel, CSV, Google Sheets).

Introducción a operaciones repetibles (automatizar flujos).

Conexión con APIs (ej: enriquecer datos con GeoNames).