SITA2020

Semana de Innovación Tecnológica en Archivos

Diciembre 10-11

DERECHOS RESERVADOS

ARCHIVO GENERAL DE LA NACIÓN COLOMBIA

SITA2020 Semana de Innovación Tecnológica en Archivos

Inteligencia Artificial y Aprendizaje Automático al servicio de los archivos

Cómo empoderar a los archiveros (y a los profesionales de la información) a través de las herramientas tecnológicas.

María Fuertes

mfuertes@libnova.com

DERECHOS RESERVADOS

La cultura es de todos Mincultura

LIBNOVA RESEARCH LABS (Contexto)

- Área de LIBNOVA dedicada a la investigación y a las relaciones con las universidades y otros centros académicos
- Objetivo: **simplificar las tareas de preservación digital** a través de la aplicación de técnicas de Inteligencia Artificial y Aprendizaje Automático.
- PROYECTO CDTI IDI – 20180287: Nuevas técnicas de catalogación mediante Inteligencia Artificial.
- ARCHIVER Project: Archiving and preservation for research environments – Grant agreement No 824516.
- Más Info: https://es.libnova.com/proyectos-idi/

DERECHOS RESERVADOS

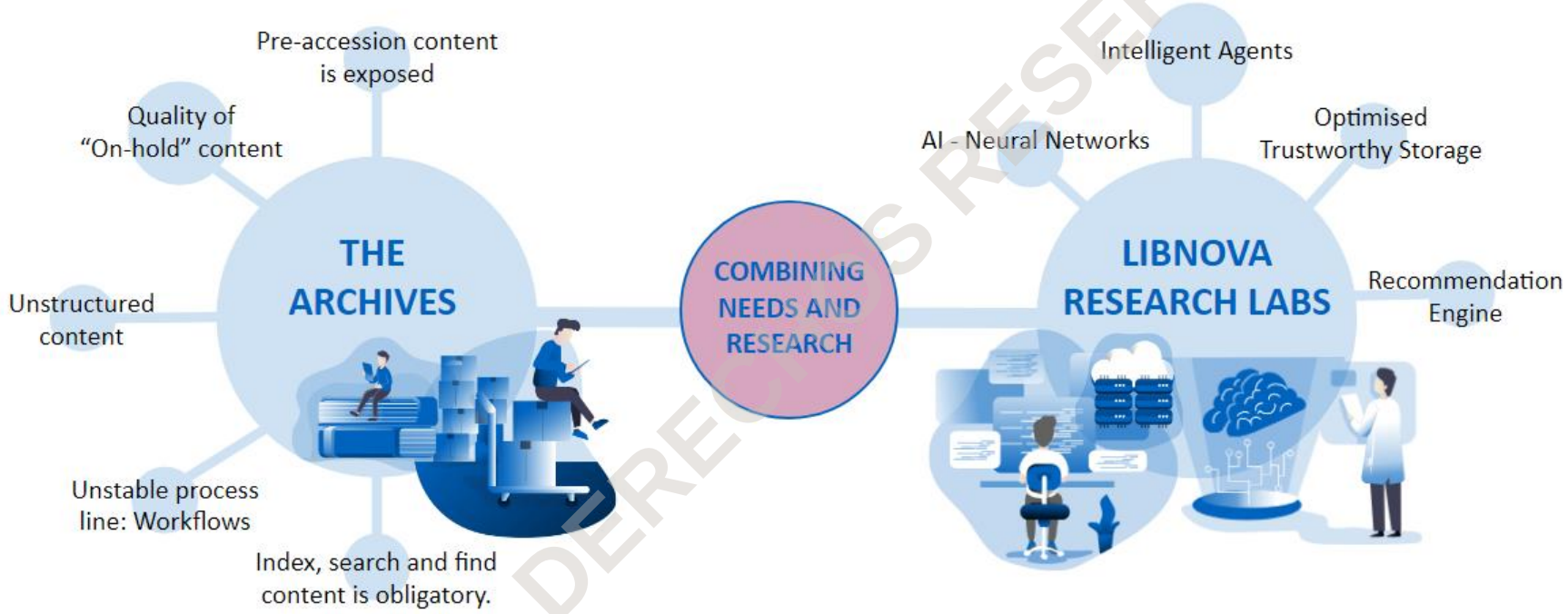
ARCHIVO GENERAL DE LA NACIÓN COLOMBIA

SITA2020 Semana de Innovación Tecnológica en Archivos

¿Qué hacemos en LIBNOVA RESEARCH LABS?

Analizamos los desafíos de los archivos…

…y desarrollamos soluciones que ayuden a los archiveros.



DERECHOS RESERVADOS

La cultura es de todos Mincultura

ARCHIVO GENERAL DE LA NACIÓN COLOMBIA

SITA2020 Semana de Innovación Tecnológica en Archivos

¿Cuál ha sido nuestro mayor hallazgo?

Inteligencia Artificial y Aprendizaje Automático al servicio de los archivos:

- Una de las cuatro áreas de innovación con las que trabajamos en el laboratorio de investigación es la aplicación de la Inteligencia Artificial y el Aprendizaje Automático al campo de la preservación digital.
- Estas son las tres funciones más interesantes que destacan los usuarios al utilizarla:
  1. Clasificación de documentos por categorías.
  2. Detección de anomalías (control de calidad).
  3. Generación automática de metadatos.

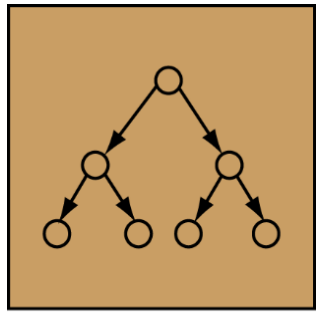
DERECHOS RESERVADOS

1. Clasificación de documentos por categorías:

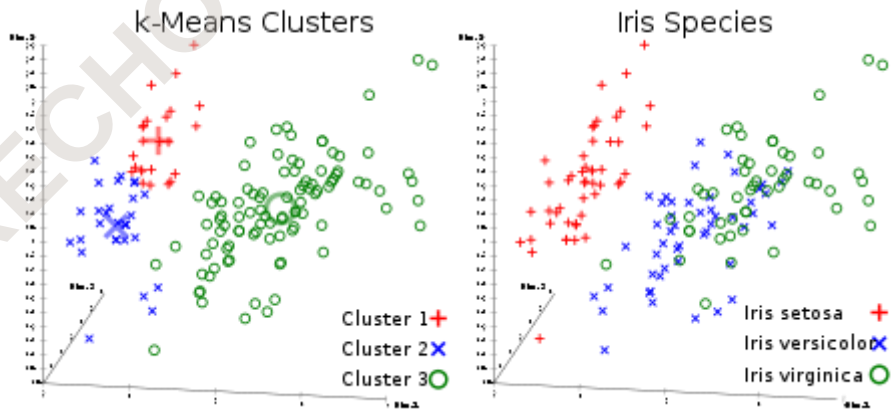
- La plataforma es capaz de detectar valores pre-organizados para hacer una clasificación en diferentes categorías.

- Ejemplos de clasificación: por organismo, por persona, por equipo, por naturaleza del documento, por fecha, etc.

- Se utilizan **Algoritmos de Clasificación Binarios** (Manuscrito o No Manuscrito) **o Multiclase** (Época).



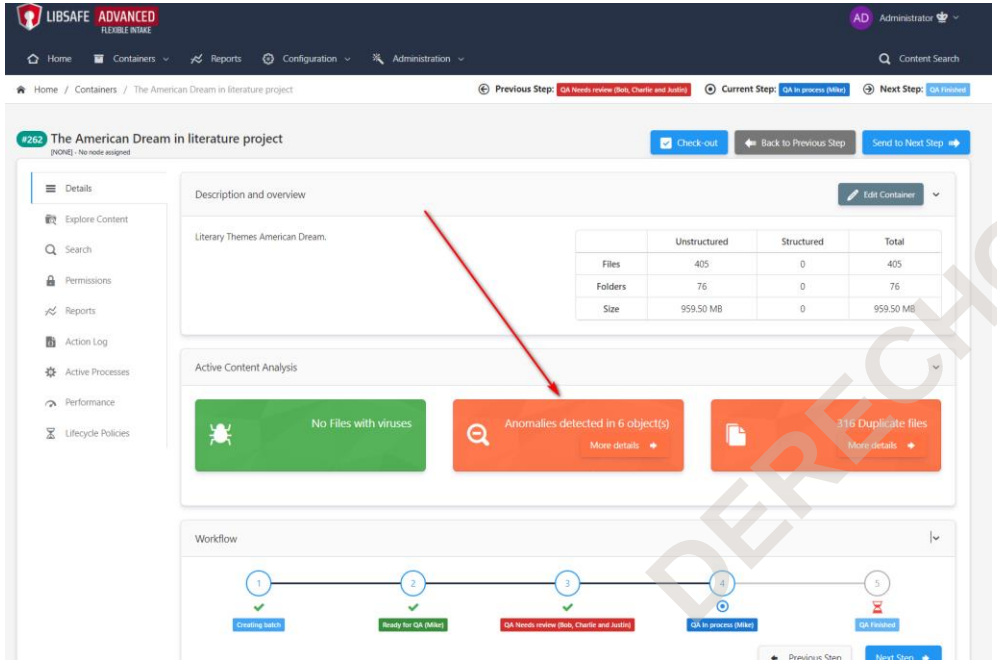
Algoritmos de
Árbol de Decisión

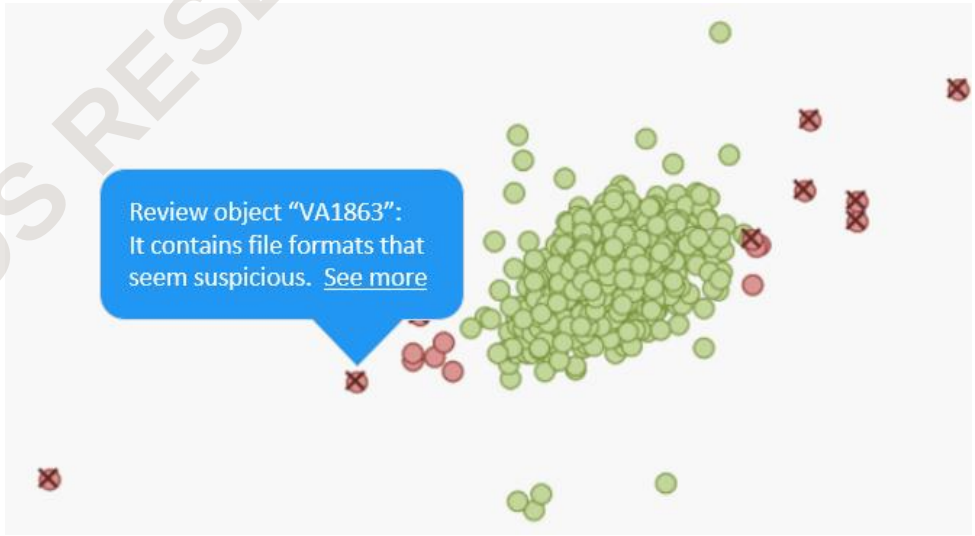


DERECHOS RESERVADOS

2. Detección de anomalías (Control de calidad):

- La plataforma es capaz de detectar valores atípicos a nivel de archivo y de objeto mediante el uso de **Autocodificadores de Redes Adversariales**. Esto ayuda al usuario a detectar objetos que no están alineados con objetos previamente vistos (outliers) que a la vez refuerza el aprendizaje de la máquina:





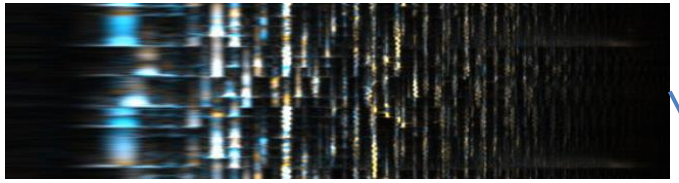
Analiza: Número de archivos, propiedades del archivo, contenido de imagen/audio/video (imagen B/N, texto/imagen, mapas, etc.), tipos de archivo, contenido de OCR, Metadatos.

DERECHOS RESERVADOS

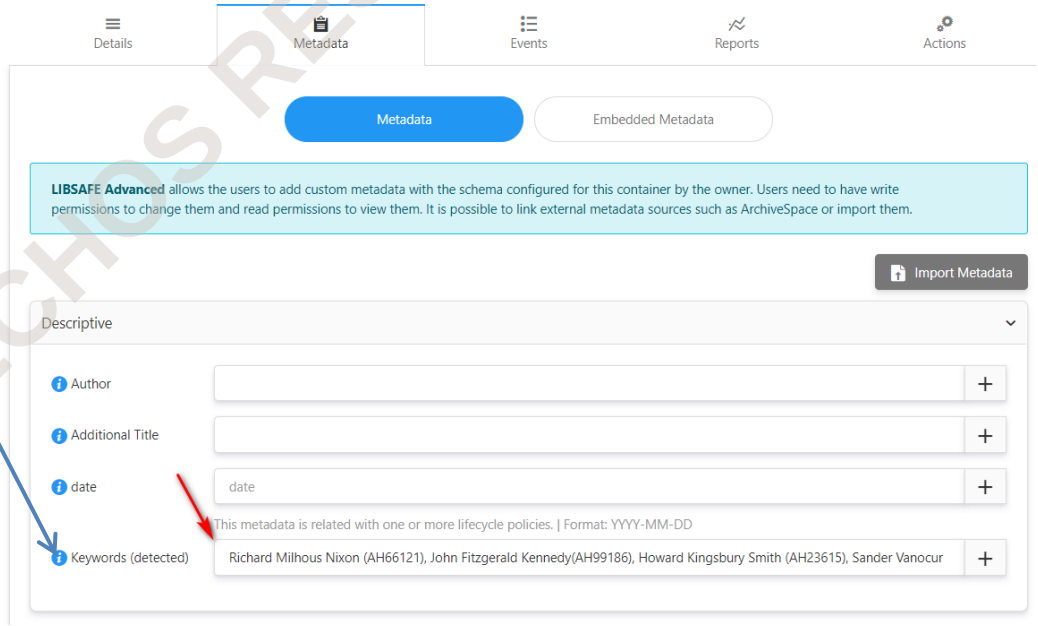
3. Generación automática de metadatos

- El software detecta quién aparece o habla en un archivo de vídeo (RNN – Recurrent Neural Network) o de audio (HMM – Hidden Markov Models) y añade sus nombres a los metadatos del objeto.

Patrones de voz:



Metadatos:



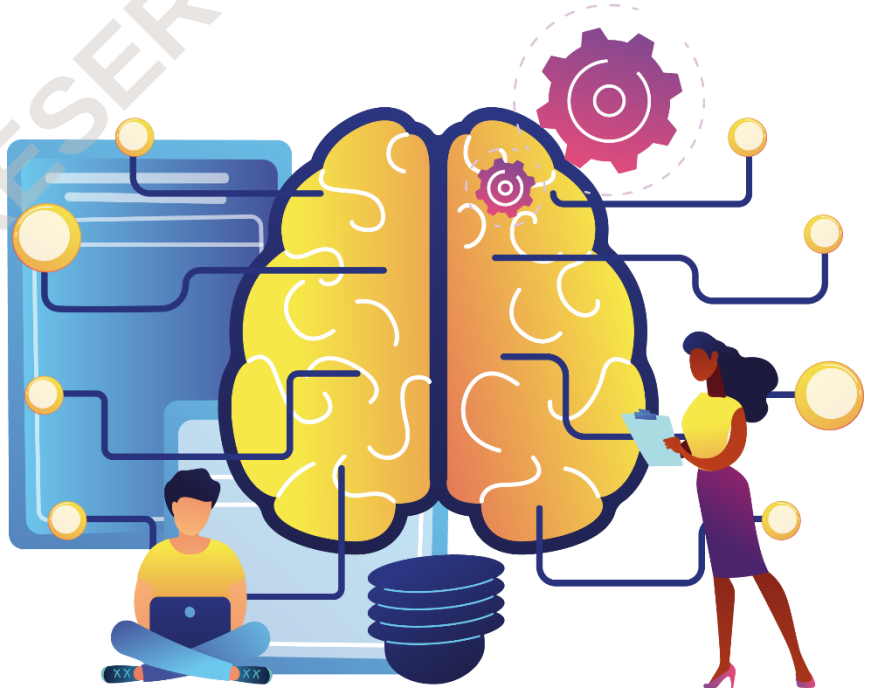
DERECHOS RESERVADOS

Inteligencia Artificial y Aprendizaje Automático al servicio de los archivos:

- Hemos visto 3 técnicas de Inteligencia Artificial:
  1. Clasificación de documentos por categorías.
  2. Detección de anomalías.
  3. Generación automática de metadatos.

- Su aplicación ayuda a los profesionales de la información a centrar sus esfuerzos en la aplicación de sus conocimientos y no perder el tiempo en cuestiones mecánicas o repetitivas.

- Por eso defendemos la idea de que las herramientas tecnológicas sirven para empoderar a los archiveros (y a los profesionales de la información).



DERECHOS RESERVADOS

SITA2020
Semana de Innovación Tecnológica en Archivos
Diciembre 10-11

GRACIAS!

María Fuertes Pozuelo
mfuertes@libnova.com
contacto@libnova.com
www.libnova.com

DERECHOS RESERVADOS