



HACKATHON RIIAA 2021

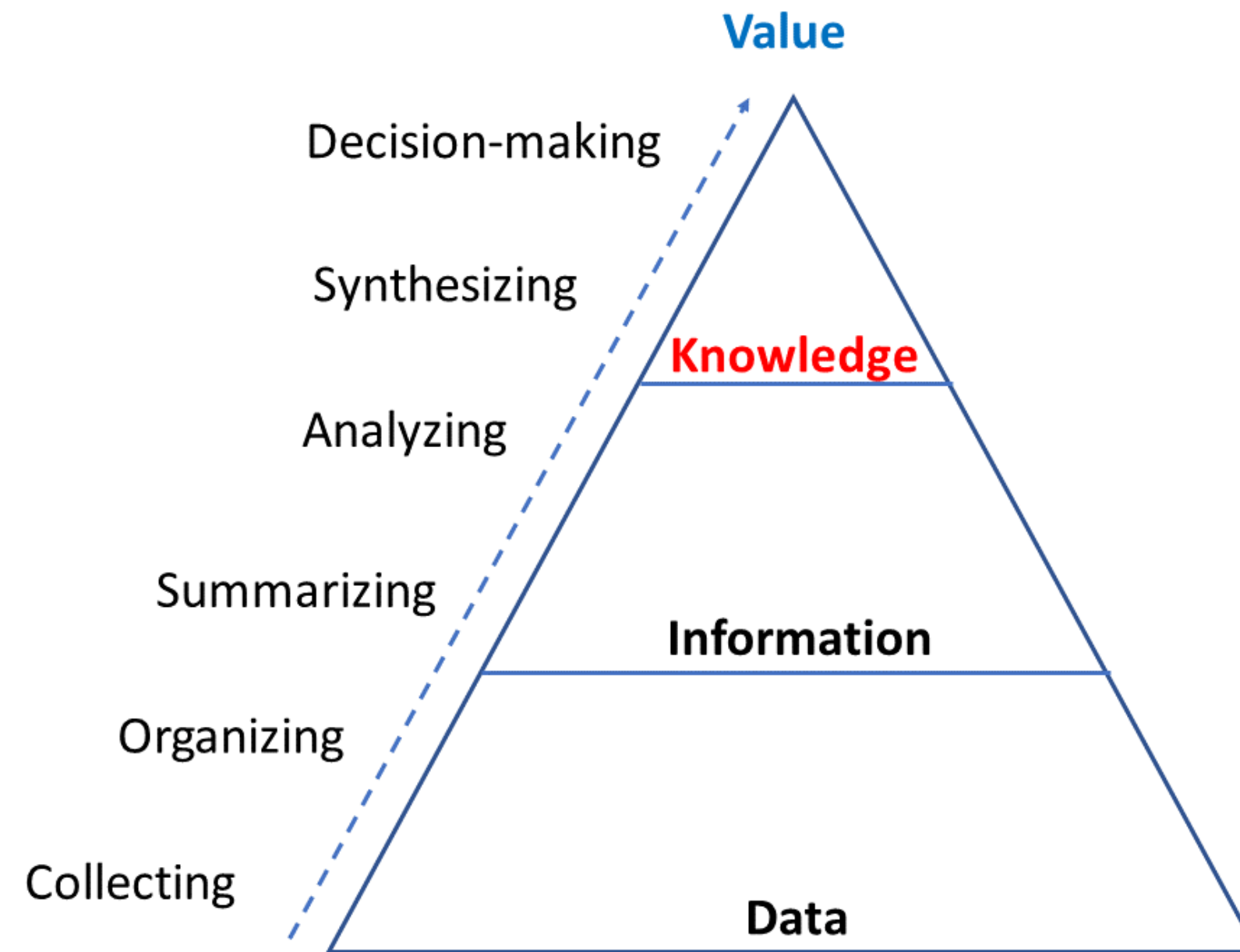
"JUSTICIA PARA LOS DESAPARECIDOS"

EQUIPO: COEUS AI

RETO 2: Extracción de texto



Problema a abordar



Propuesta de valor

Coeus AI es un sistema robusto y flexible para el análisis inteligente de documentos.

Las funcionalidades incluyen:

- Tratamiento de imágenes
- Extracción de texto
- Un motor de búsqueda semántico capaz de comprender consultas en lenguaje natural y retornar los documentos más adecuados.

Planteamiento de la solución



Extracción de expedientes
Extracción de entidades



Pre- procesamiento de documentos

Extracción automatica de texto

Post- procesamiento de documentos

Integración en un motor de búsqueda semantico



Solución Local



Solución en la nube

Evaluación de texto
Mediante métricas

Motor OCR	Pre-procesamiento	Dice	Jaccard	Coseno
Azure	Otsu	0.853	0.813	0.853
Azure	Adaptive	0.857	0.82	0.857
Azure	No	0.768	0.74	0.768

Coeus AI

Tratamiento automático de imágenes:

- Eliminación de sombras
- Eliminación de ruido
- Binarización adaptativa
- Binarización de Otsu
- Corrección de orientación



Tesseract OCR

Extracción de texto con Azure CS



Extracción de texto con EasyOCR



Evaluación de texto:

- Szigriszt-Pazos
- Gutierrez-Polini
- Language Tool

spaCy + 

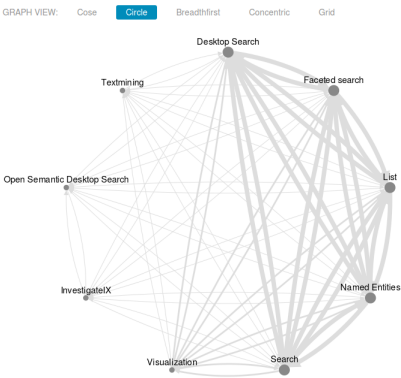
- Fechas con DateParser
- Expedientes con Regex

• Texto evaluado

• Imágenes procesadas

Unificación e indexado:

- Open Semantic Search

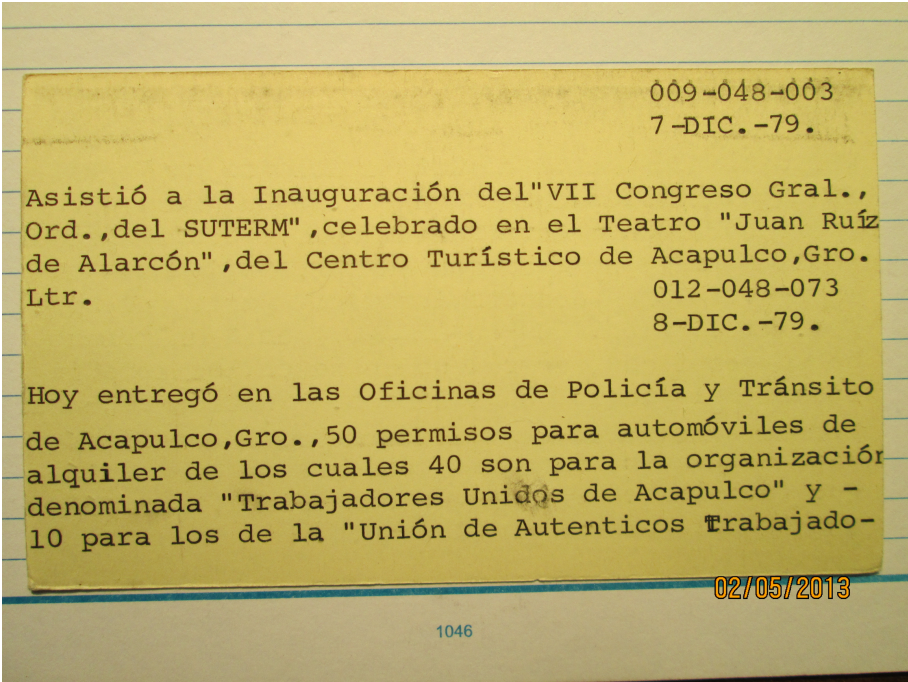


Tratamiento de Imágenes

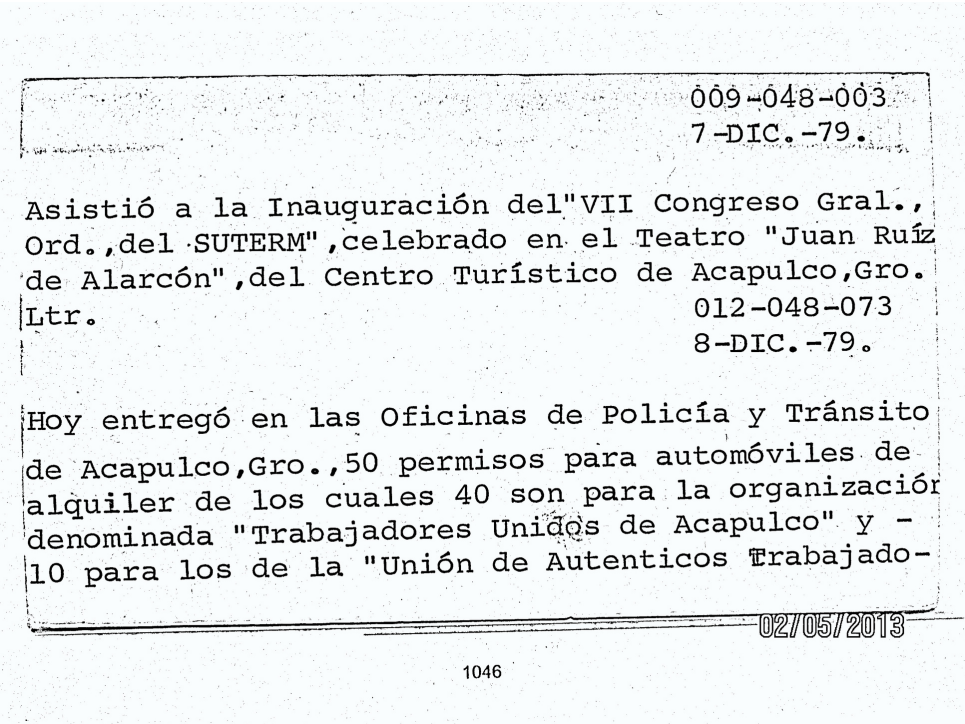
Demostración



ANTES



DESPUÉS



Search

Search options

ListPreviewEntibesImagesVideosAudiosMapAnalyze

Sort

Relevance

PreviousPage 1 of 1 (results 1 to 2 of 2)Next

Ficheros_VARP_VARP_relaciones_IMG_5856.txt

2021-08-23T03:12:31Z

Ficheros_VARP_VARP_relaciones_IMG_5856.txt

• instalaciones de la Preparatoria 0.7, "Salvador Allen- de ,en Acapulco, ro. , y tuvo la intención de que sus — simpatizantes se posesionaren de los predios que exis- tian en la Barranca de La baja pero LEANDRO VELEZ — ABARCA y RUFINA CASTELLANOS PENA se adelantaron a la - acción y se

Persons:LEANDRO VELEZRUFINA CASTELLANOSHORTENCIA SANTOYO NUNEZOrganizations:Sindico Proc. Exp-100-10-1-75 U.A.G. Locations: Ayunt. Escuelas Acapulco Gro Barranca Plaza Principal Barranca de La baja "a Plaza Principal Salvador Allen Phone numbers: 100-10-1-75

Open | Tagging & annotation | Preview

Ficheros_VARP_VARP_relaciones_IMG_5856.JPG

2021-08-22T15:16:27Z

Ficheros_VARP_VARP_relaciones_IMG_5856.JPG

• instalaciones. de da Preparatoria | "We «7, "Salvador Allen de. ¿en Acapul@, ros. y tuvo la intención de que sus main | simpatizantes. se posesionaren de los. predios que exis- | - tian en la Barranca de La Laja pero LEANDRO VELEZ === ABARCA y RUFINA CASTELLANOS. PEÑA se

Persons:Salvador AllenSindico Proc. Lider RUFINA CASTELLANOS Allen LEANDRO VELEZOrganizations:— - ae: A. G. Sindico Proc.,del Ayunt.,de Acapulco, A Locations: @a Plaza Principal Barranca de La Laja Acapulco Plaza Principal Ayunt. Acapul@ ¿ " Barranca Gro Phone numbers: 100-10-1-75

Open | Tagging & annotation | Preview

PreviousPage 1 of 1 (results 1 to 2 of 2)Next

File date

2021 (2)

Path(s)

media (2) -

Show less (-) | more (+)

Persons

LEANDRO VELEZ (2) -

Allen (1) -

HORTENCIA SANTOYO NUNEZ (1) -

Lider (1) -

RUFINA CASTELLANOS (1) -

RUFINA CASTELLANOS. (1) -

Salvador Allen (1) -

Sindico Proc (1) -

Show less (-) | more (+)

Organizations

Exp-100-10-1-75 (1) -

Sindico Proc. (1) -

Sindico Proc.,del Ayunt.,de

Acapulco, A (1) -

U.A.G. (1) -

— - ae: A. G. (1) -

Show less (-) | more (+)

Locations

Acapulco (2) -

Ayunt (2) -

Barranca (2) -

Gro (2) -

Coeus AI

JusticIA

coeus-ai.ddns.net/search/