

Vigilancia e Inteligencia en redes sociales y medios digitales

Carlos Villalobos Cascante

Marcos Chavarría Jiménez

Carlos Sabat Zamora

10 de noviembre del 2022

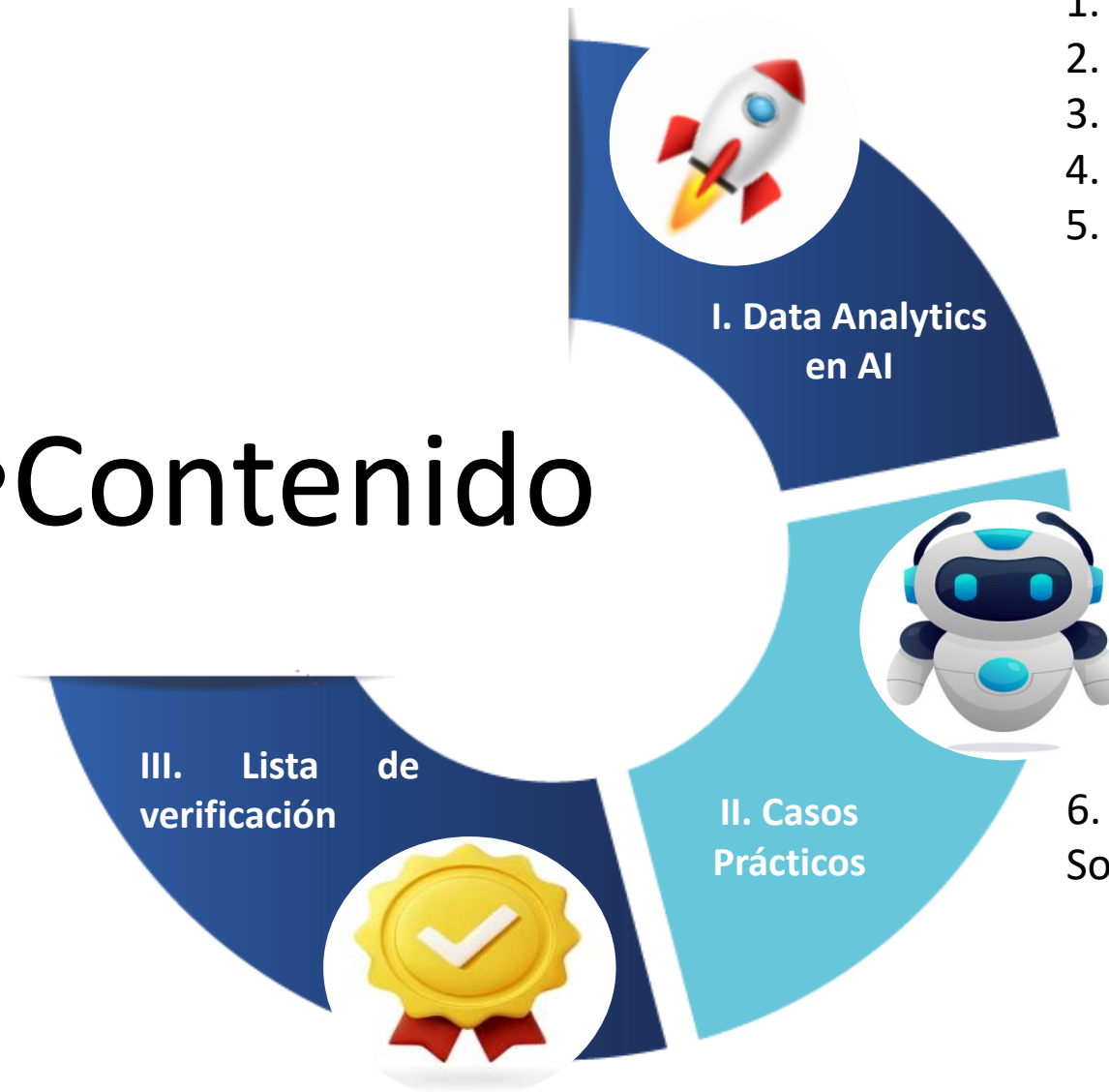
Jornada
2022

 **ISACA.**
Costa Rica Chapter

 Instituto de
Auditores Internos
de Costa Rica

1. Introducción
2. Data Analytics
3. Científico de Datos
4. Beneficios
5. Tipos de Datos

•Contenido



6. Vigilancia e Inteligencia Redes Sociales y Medios Digitales

1. Introducción

Jornada
2022

El **papel de la auditoría** interna se está transformado debido a la **complejidad** de las actividades comerciales y a los **nuevos retos** en un mundo volátil. Nuestra profesión hoy no solo debe respetar el cumplimiento de las normativas, sino también **reducir los costos** y proporcionar información detallada de nuestros interesados como **consultor estratégico**.



El poder de la tecnología cognitiva y cómo está revolucionando el proceso de auditoría de forma **exponencial**. Los profundos cambios que tienen lugar en el espacio de la auditoría a medida que evoluciona la tecnología cognitiva, combinados con otros desarrollos innovadores, como la automatización de procesos robóticos (RPA) y el análisis avanzado, está cambiando el panorama de la auditoría de **forma permanente**.



I. Data Analytics en Auditorías Internas

Jornada
2022

 **ISACA.**
Costa Rica Chapter

 Instituto de
Auditores Internos
de Costa Rica

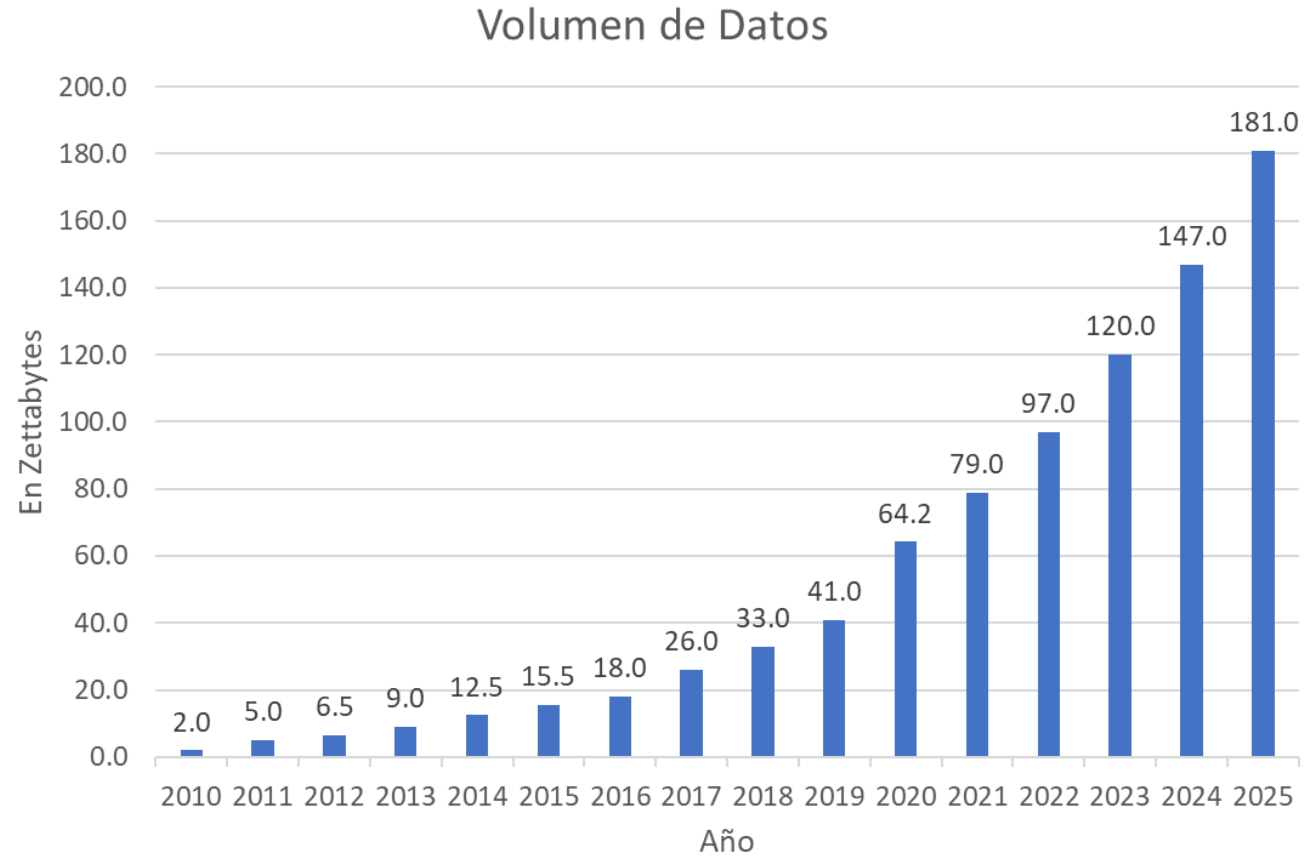
2. Data Analytics en Auditorías Internas

En un mundo cada vez más hiperconectado, en donde el cambio es la única constante, las organizaciones deben ser permeables y desarrollar una capacidad de adaptabilidad para lograr el éxito.

La globalización, la transformación digital y las diferentes disrupciones han modificado los modelos tradicionales de negocios, así como las expectativas de los clientes. Nuestra profesión no escapa de esta revolución.

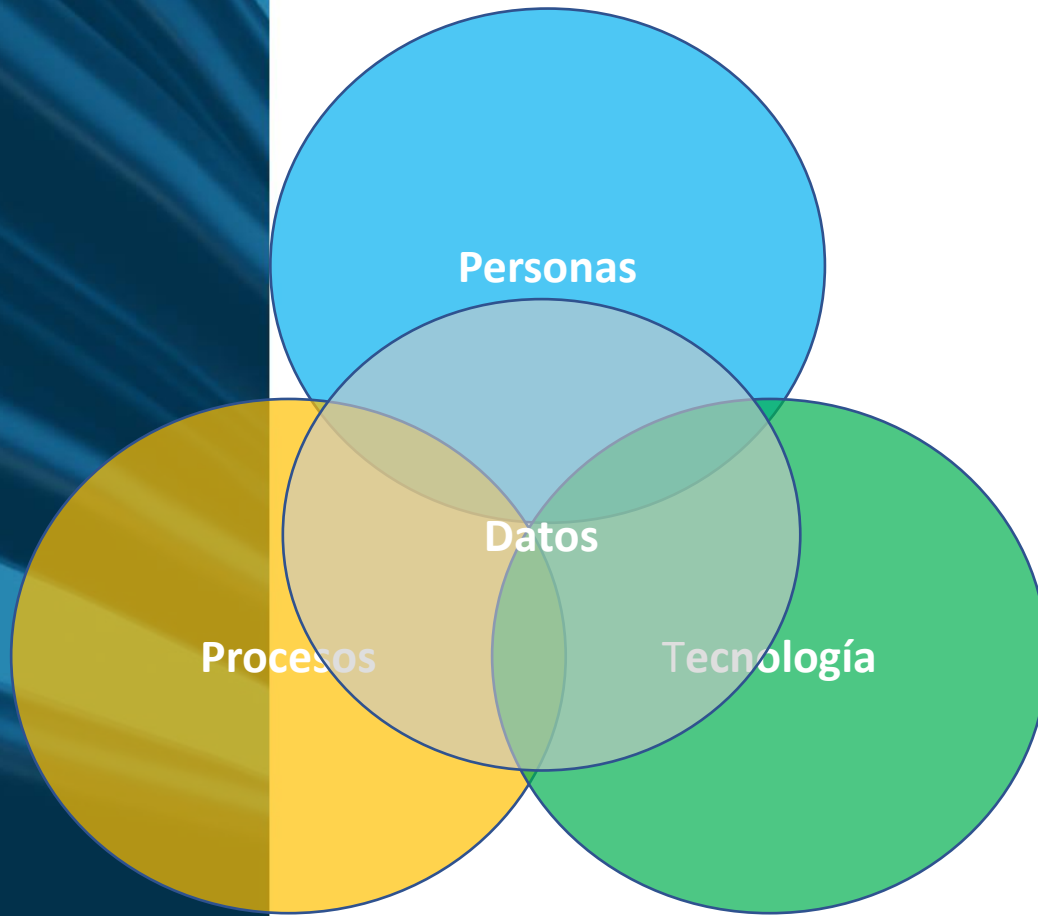
Esta revolución y los riesgos que conlleva, el tomar la decisión correcta y en el momento adecuado se convierte en una obligación, por ello, las Auditorías deben incorporar la analítica de datos en cada una de sus fases del proceso de auditoría, para poder responder con eficiencia, ante nuevas exigencias de nuestros clientes en un entorno mucho más volátil, incierto y complejo.

Datos del Departamento de Investigación de Statista, **la cantidad de datos** creados, capturados, copiados y consumidos a nivel mundial **crecen de manera acelerada**



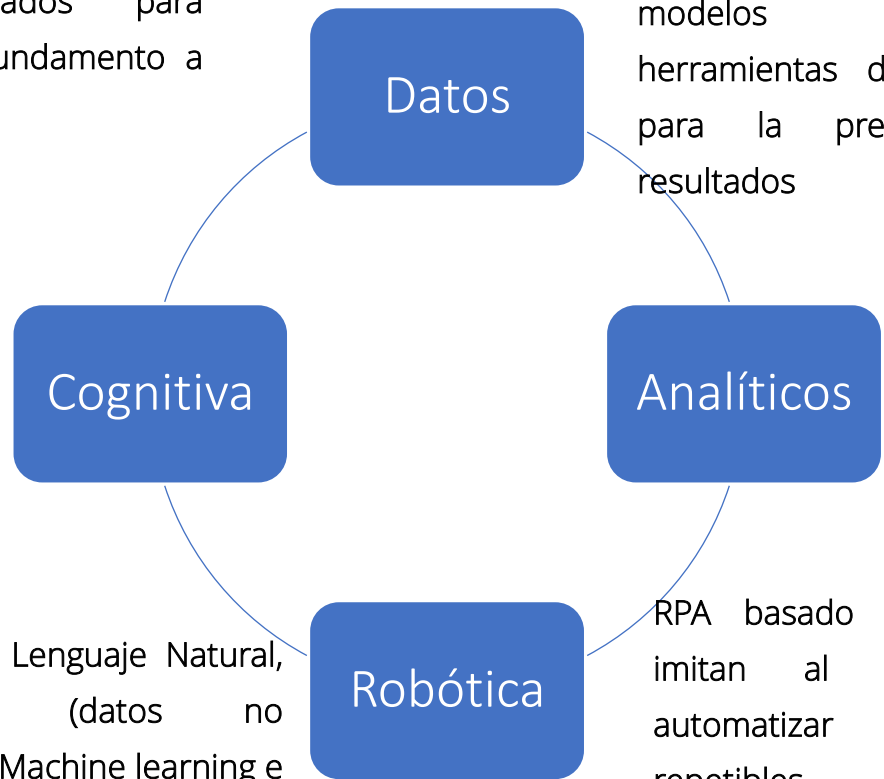
1 ZB = 1m TB

Tendencias era digital



Datos Integrados para proporcionar fundamento a la información

Soluciones para desarrollar modelos predictivos, herramientas de visualización para la presentación de resultados



Generación de Lenguaje Natural, procesamiento (datos no estructurados), Machine learning e inteligencia artificial

RPA basado en reglas que imitan al humano para automatizar procesos repetibles

3. ¿Qué es un Científico de Datos?

"Persona mejor en **estadística** que cualquier desarrollador, y mejor en **programación** que cualquier estadístico", Josh Wills, Director de Ingeniería de Datos en Slack.

"Es una persona **con fundamentos en matemáticas, estadística y métodos de optimización**, con conocimientos en **lenguajes de programación** y que además tiene una **experiencia práctica** en el análisis de datos reales y la elaboración de **modelos predictivos**" José Antonio Guerrero. Científico de datos de España

De acuerdo a Deloitte, "un científico de datos es el **profesional** que, ante enormes bases de datos, la mayor parte de ellos no estructurados aplica sobre ellas sus conocimientos en **programación, matemáticas y estadística** para recopilar, extraer, procesar **información relevante** que contienen".



Cualidades



Creatividad

Utilizar la imaginación para resolver incógnitas en la comprensión de datos



Analítico

Capacidad de síntesis, de relacionar, generar conexiones, aplicar metodología



Comunicador

Transmitir de forma eficaz qué ha pasado y por qué

Equipos Multidisciplinarios y ágiles

4. Beneficios



5. Tipos de Datos

Estructurados



A central illustration on a light blue background. At the top left is a white grid icon. In the center is a computer monitor displaying a webpage with horizontal lines. To the left of the monitor are three interlocking gears. To the right is a line graph with three data points and an upward-pointing arrow. Below the monitor is a white database cylinder icon. A white arrow points upwards from the bottom left towards the monitor.



VS



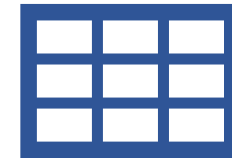
No Estructurados

Datos Estructurados

Son aquellos que se encuentran en la mayoría de las bases de datos, tienen un **formato estandarizado**, una estructura bien definida que se suelen mostrar en filas y columnas, cumplen con un modelo de datos, siguen un orden persistente y son de **fácil acceso para humanos y programas**.

Los ejemplos de datos estructurados:

- Hojas de cálculo
- Bases de Datos relacionales
- Aplicaciones de cuestionarios en tipo Test
- Formularios web
- Encuestas digitales a usuarios de servicios



Ejemplo uso de datos estructurados

RPA WEB Públicas de entidades como SUGEF y BCCR

Entidad	Cuenta	Saldo	Debe/Haber
BCCR - Banco Nacional	100-00-000	1.234.567.890,12	Debe
BCCR - Banco Nacional	100-00-001	987.654.321,01	Haber
BCCR - Banco Nacional	100-00-002	543.210.987,65	Debe
BCCR - Banco Nacional	100-00-003	210.987.654,32	Haber
BCCR - Banco Nacional	100-00-004	876.543.210,98	Debe
BCCR - Banco Nacional	100-00-005	432.109.876,54	Haber
BCCR - Banco Nacional	100-00-006	109.876.543,21	Debe
BCCR - Banco Nacional	100-00-007	765.432.109,87	Haber
BCCR - Banco Nacional	100-00-008	321.098.765,43	Debe
BCCR - Banco Nacional	100-00-009	987.654.321,01	Haber
BCCR - Banco Nacional	100-00-010	543.210.987,65	Debe

DB Tabla Dinamica



Cuenta	Saldo	Debe/Haber
100-00-000	1.234.567.890,12	Debe
100-00-001	987.654.321,01	Haber
100-00-002	543.210.987,65	Debe
100-00-003	210.987.654,32	Haber
100-00-004	876.543.210,98	Debe
100-00-005	432.109.876,54	Haber
100-00-006	109.876.543,21	Debe
100-00-007	765.432.109,87	Haber
100-00-008	321.098.765,43	Debe
100-00-009	987.654.321,01	Haber
100-00-010	543.210.987,65	Debe

BalanceSituacion



Activo



Entidad	Activo	Pasivo
BCCR	1.234.567.890,12	987.654.321,01
Banco de Costa Rica	987.654.321,01	543.210.987,65
Banco de Puntarenas	543.210.987,65	210.987.654,32
Banco de Occidente	210.987.654,32	876.543.210,98
Banco de Guayaquil	876.543.210,98	432.109.876,54
Banco de Cuenca	432.109.876,54	109.876.543,21
Banco de Loja	109.876.543,21	765.432.109,87
Banco de Esmeraldas	765.432.109,87	321.098.765,43
Banco de Bolívar	987.654.321,01	987.654.321,01
Banco de Santo Domingo	543.210.987,65	543.210.987,65

Comparativo Bancos



Producción y Empleo



Indices de Precios



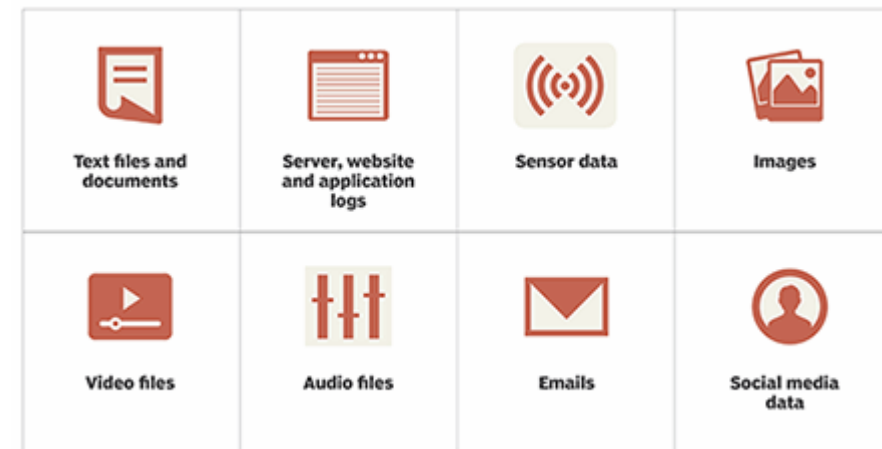
Datos no estructurados

Los datos no estructurados son datos que **carecen** de una **estructura** o arquitectura identificable. Generalmente son datos binarios que no tienen estructura interna. Corresponde a un cúmulo masivo y desorganizado de varios objetos que no tienen valor hasta que se logran estructurar.

Este tipo de datos supone alrededor de un **80% de todo el existente** en cualquier organización y su manejo es más complejo que los estructurados. Su transformación es costosa y compleja.

Los ejemplos de datos no estructurados son:

- Correos electrónicos
- Archivos o imágenes en PDF
- Imágenes Digitales
- Videos
- Redes Sociales
- Medios Digitales
- Audio, otros

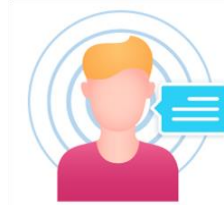


Servicios Cognitivos

- Servicios en la nube.
- Pone la **inteligencia artificial** al alcance de los desarrolladores y científicos de datos.
- Le dan un **lado humano a las tecnologías**, para insertar la capacidad de **ver, escuchar, hablar, buscar, comprender y acelerar la toma de decisiones avanzada en las aplicaciones.**



Vista



Habla



Lenguaje



Búsqueda

Características

Proyecto

- Detección de objetos, escenas y actividades.
- Reconocimiento e identificación facial.
- Reconocimiento de celebridades y hitos.
- **Reconocimiento de emociones.**
- **Reconocimiento de texto y escritura a mano (OCR).**
- Reconocimiento de imagen personalizable.
- Metadatos de vídeo, audio y extracción y análisis de fotogramas clave.
- Moderación de contenido explícita u ofensiva .

- **Transcripción del habla (voz a texto).**
- **Modelos de voz personalizados para vocabularios únicos o entornos complejos.**
- **Texto a voz.**
- Voz personalizada.
- Traducción de voz en tiempo real.
- Transcripción y traducción de voz personalizables.
- **Identificación y verificación de altavoces.**

- **Detección de idioma.**
- **Reconocimiento de entidades con nombre.**
- **Extracción de frases clave.**
- **Análisis de sentimientos de texto.**
- Revisión ortográfica plurilingüe y contextual.
- Moderación del contenido de texto explícito u ofensivo.
- Traducción de texto.
- Comprensión del lenguaje contextual.
- **Extracción de texto no estructurado**
- **Creación de la Base de Conocimiento.**

- **Resultados de búsqueda web, noticias, imágenes y vídeos sin publicidad.**
- **Tendencias de vídeo, noticias.**
- Identificación de imágenes, clasificación y extracción de conocimientos.
- Identificación de imágenes y productos similares.
- Reconocimiento y clasificación de entidades con nombre.
- Adquisición de conocimientos para entidades nombradas.
- Creación de motores de búsqueda personalizados sin publicidad.

EVONY

TRANSCRIPTOR
DE ACTAS

PÉRDIDA DE CLIENTES
VIGILANCIA DE MEDIOS

PÉRDIDA DE CLIENTES
VIGILANCIA DE MEDIOS

II. Caso Práctico: Redes Sociales y Medios Digitales



VI sobre Redes y Medios Digitales

Jornada
2022

¿Para qué usan las
RRSS y MD en la
Entidad?

¿Quiénes gestionan
las RRSS y MD en la
Organización?

Estrategia de
Comunicación y
RRSS: Cliente &
Reputación & Marca

Diseñar un modelo analítico de escucha activa, soportado en ciencia de datos, que permita crear una herramienta de aseguramiento sobre redes sociales y medios digitales.

Redes y Medios Canal - Comunicación - **Marca** - Reputación - **Cliente**

Posicionamiento

Percepción

Partes Interesadas

Prototipo

Jornada
2022

Tablero Gerencial
Cuadro de Mando
Integral



IRR RRSS (AG)

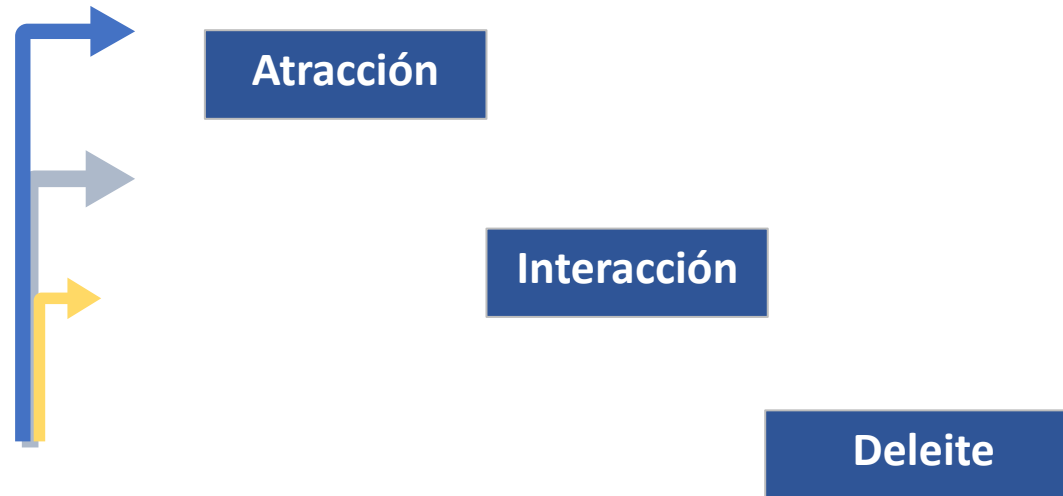


I. Marca RRSS (AG)



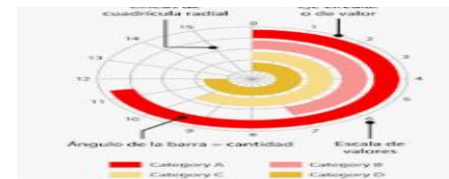
I. Benchmarking RRSS (AG)

Tablero de
Seguimiento

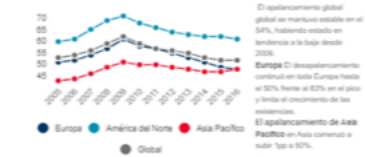


Informe de
Contenido
& Reporteria

	f	ig	tw	yt	in
Positivas	15%	60%	27%	80%	30%
Neutras	20%	35%	73%	15%	50%
Negativas	65%	25%	0%	5%	20%



Apalancamiento (Deuda como % de acciones)



ISACA
Costa Rica Chapter

Instituto de
Audidores Internos
de Costa Rica



Técnicas de extracción de datos

- Extracción de datos de la web (Medios, redes sociales):
 - WebScraping
 - APIs
 - Limpieza y Expresiones regulares



Extracción de datos de la web

1. WebScraping

Definición

- Consiste en extraer datos de sitios web de manera automatizada.
- El scraping de datos implica localizar los datos y luego extraerlos. No copia y pega, sino que, obtiene directamente los datos de manera precisa. No se limita a la web, los datos se pueden obtener prácticamente desde cualquier lugar donde se almacenan. Puede ser Internet u otra fuente de datos.



Ventajas	Retos
Extracción de datos de forma automatizada	Estructura de datos
Disminuir cargas de trabajo	Limpieza y ordenamiento de datos
Manejo de grandes cantidades de datos	
Aumento en la velocidad de procesos	

Técnica WebScraping

[link](#)



Jornada
2022

El Observador

Pymes reportan una mayor mora crediticia en el Banco Nacional que antes

Por JOSUÉ ALVARADO · 12:08 AM 22 SEPTIEMBRE, 2022

time.entry-date.updated.td-module-date 157.57 x 14.67

```
...  
<time class="entry-date updated td-module-date" datetime="2022-09-22T00:08:15+00:00">12:08 AM 22 septiembre, 2022</time>  
== $0  
</div>  
</div>  
</div>  
</div>  
</div>
```

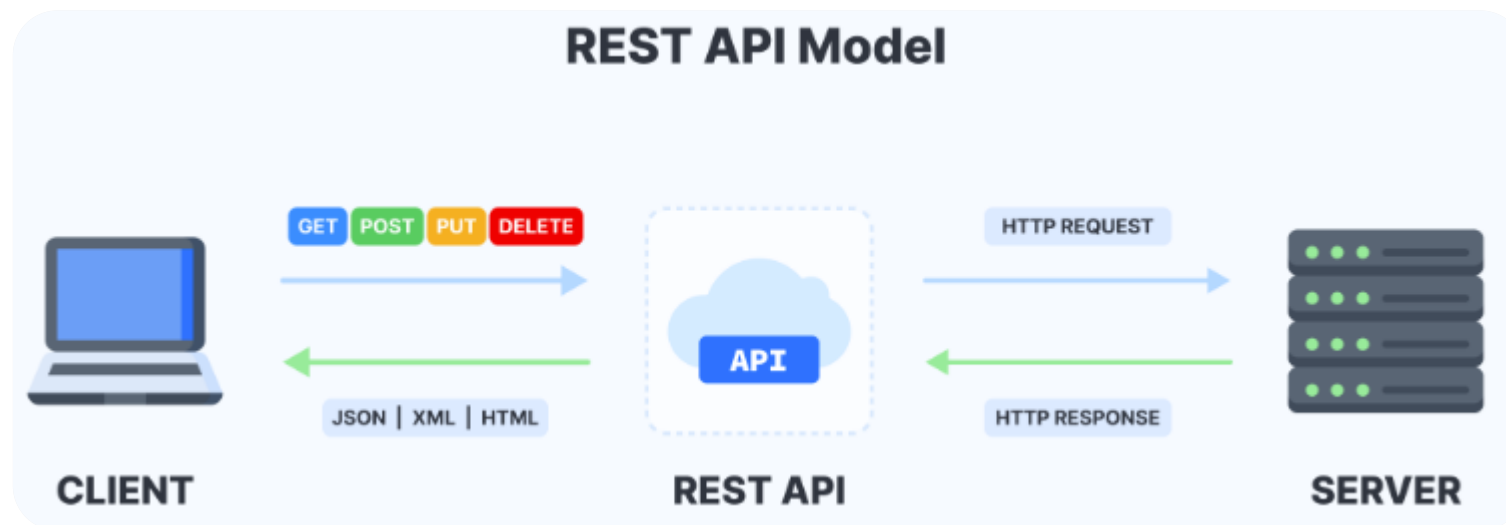
```
from datetime import datetime  
from bs4 import BeautifulSoup  
import pandas as pd  
from pyarsing import col  
from selenium.webdriver.common import by
```

```
#Obtiene la fecha de la noticia  
  
titulo_fecha = soup.find_all("time", {"class": "entry-date updated td-module-date"})  
if (len(titulo_fecha)>0):  
    titulo_fecha = titulo_fecha[6].text  
else:  
    titulo_fecha = ""
```

Extracción de datos de la web

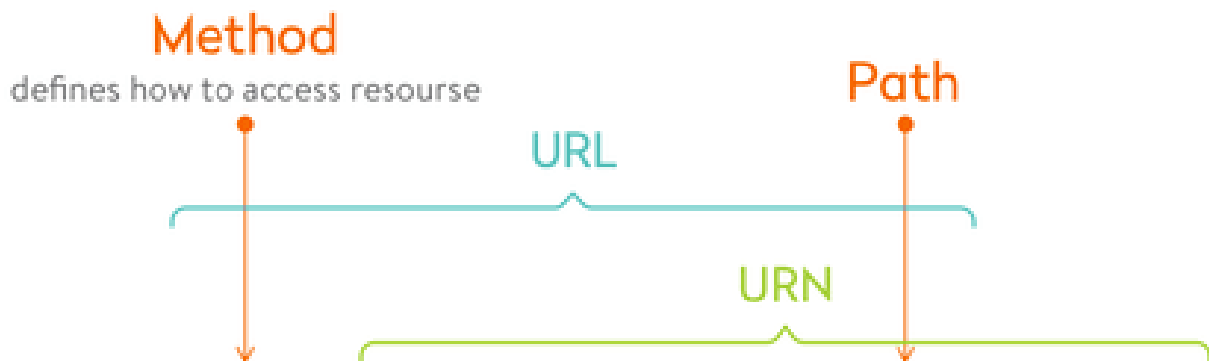
2. Interfaz de programación de aplicaciones (API)

La interfaz de programación de aplicaciones es un conjunto de subrutinas, funciones y procedimientos o métodos, en la programación orientada a objetos, que ofrece cierta **biblioteca de servicios** para ser utilizada por otro software como una **capa de abstracción**.

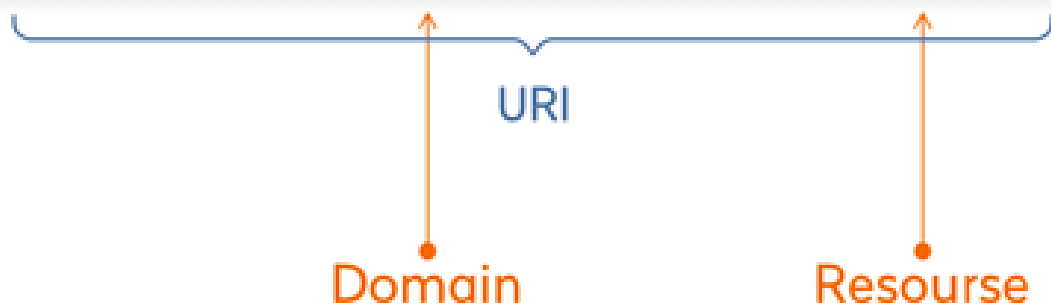


Acceso y seguridad APIs

Jornada
2022

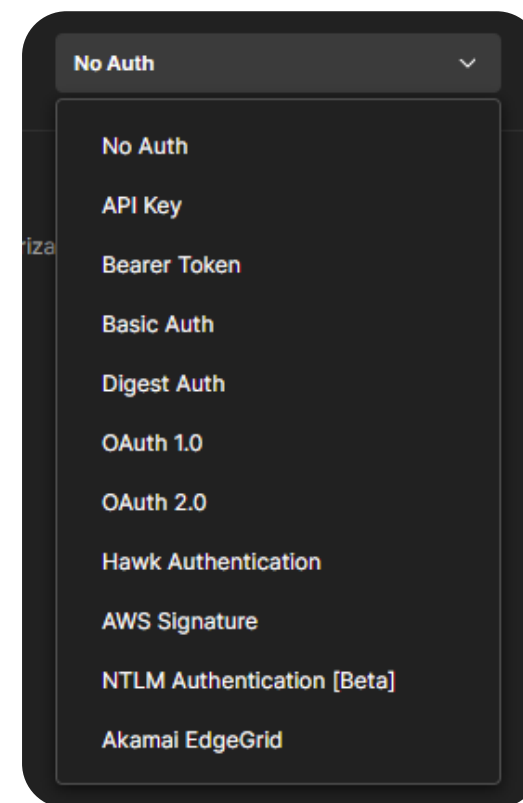


https://examplesite.com/page/example



[Components of URI]

🔑 Authorization method



Ejemplo de token: <https://jwt.io>

← → ↻ 🏠 developers.facebook.com/docs/graph-api/get-started

Meta for Developers Docs Tools Support My Apps 🔍 Search developer documentation

Graph API
 Overview
Get Started
 Graph Explorer Guide
 Guides
 Server-Sent Events
 Changelog
 Features Reference
 Permissions Reference
 Reference

Get Started

This guide explains how to get started with receiving data from the Facebook Social Graph.

Before You Start

You will need:

- Register as a Facebook Developer
- A Facebook App – This app will be for testing so there is no need to involve your app code when creating this Facebook App.
- The Graph Explorer tool open in a separate browser window
- A brief understanding of the structure of the Facebook Social Graph from our Graph API Overview guide



Jornada
2022

<https://developers.facebook.com/docs/graph-api/get-started>

← → ↻ 🏠 developers.facebook.com/tools/explorer/1067271583999008/?method=GET&path=me%3Ffields%3Did%2Cname%2Cposts&version=v15.0

Meta for Developers Docs Tools Support My Apps 🔍 Search developer documentation 🔔 Marcos Chavar...

Graph API Explorer

GET → / v15.0 / me?fields=id,name,posts Submit

Node: me

- id
- name
- posts
- + Search for a field
- + Search for a field

```
{
  "id": "182346165134456",
  "name": "Banco Nacional",
  "posts": {
    "data": [
      {
        "created_time": "2022-11-01T20:01:02+0000",
        "message": "La prevención y educación son las armas más importantes que tenemos c
      },
      {
        "created_time": "2022-11-01T18:10:02+0000",
        "message": "Pagá tu marchamo con el BN y participá por un Toyota Corolla Cross h:
        "id": "182346165134456_5631875890181429"
      },
      {
        "created_time": "2022-11-01T17:42:15+0000",
        "message": "👤 Te subís al bus.
      },
      {
        "created_time": "2022-11-01T17:42:15+0000",
        "message": "👤 Te subís al bus.
      },
      {
        "created_time": "2022-11-01T17:42:15+0000",
        "message": "Pagás sin contacto pasando tu tarjeta por el validador.
      },
      {
        "created_time": "2022-11-01T17:42:15+0000",
        "message": "En menos de un segundo, sonará el \"beep\" de pago exitoso.
      },
      {
        "created_time": "2022-11-01T17:42:15+0000",
        "message": "¡Y listo! Podrás viajar de forma segura, rápida y ágil.
      },
      {
        "created_time": "2022-11-01T17:42:15+0000",
        "message": "La Directora de Negocios Digitales a.i., Debbie Calvo Orozco, destaca la recién incorpor
        "id": "182346165134456_5631566880212330"
      }
    ]
  }
}
```

Access Token

EAAPKrYel r5entbHGr

Generate Access Token

Meta App

Auditoria Interna

User or Page

User Token

Permissions

- × user_posts
- × read_insights
- public_profile

Add a Permission

2 options selected

<https://developers.facebook.com/tools/explorer>





<https://developers.facebook.com/docs/graph-api/reference/post/>
<https://developers.facebook.com/docs/graph-api/reference/v15.0/insights>

```
import facebook

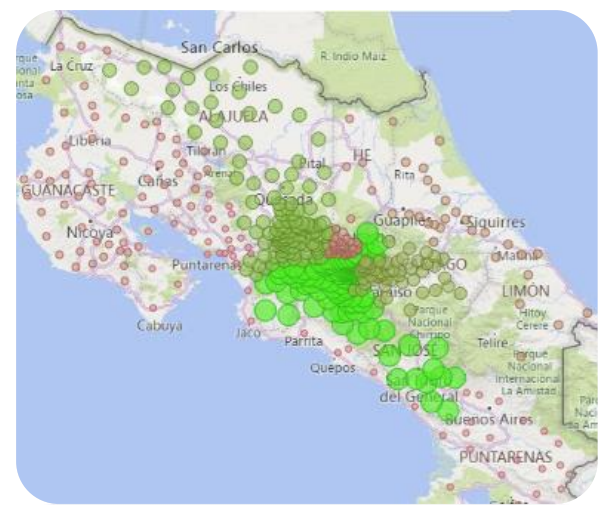
token='EAAPKrYekpCABAGgtbK4PofNh9pmw5nS8NEBgAokelRzAwI5entbHGra
graph = facebook.GraphAPI(access_token=token, version="3.1")

data = graph.get_object("/me?fields=id,name") #posts
print(data)
print(data['id'])
print(data['name'])

#post_limit=1
#data_post = graph.get_object("/me/posts?fields=id,reactions.l
#for item in data_post['data']:
#    print(item)
```

✓ 0.5s

```
{'id': '182346165134456', 'name': 'Banco Nacional'}
182346165134456
Banco Nacional
```



	profile_id	metric_id	ty...	desc	value	date
1	182346165134456	page_fans_city	day	Limón, Provincia de Limón, Costa Rica	6027	2022-09-11
2	182346165134456	page_fans_city	day	Desamparados, San José Province, Costa Rica	6268	2022-09-11
3	182346165134456	page_fans_city	day	Liberia, Provincia de Guanacaste, Costa Rica	6571	2022-09-11
4	182346165134456	page_fans_city	day	La Unión, Cartago Province, Costa Rica	6677	2022-09-11
5	182346165134456	page_fans_city	day	Guápiles, Limón Province, Costa Rica	6911	2022-09-11
6	182346165134456	page_fans_city	day	Alajuela, Provincia de Alajuela, Costa Rica	13759	2022-09-11
7	182346165134456	page_fans_city	day	Heredia, Provincia de Heredia, Costa Rica	15306	2022-09-11
8	182346165134456	page_fans_city	day	Cartago, Provincia de Cartago, Costa Rica	21588	2022-09-11
9	182346165134456	page_fans_city	day	San José, Provincia de San José, Costa Rica	55706	2022-09-11



Comparativo de técnicas

Jornada
2022



Web scraping

Raspado web

Alternativo

Mecanismo para extraer información de un sitio web **simulando la navegación.**

No involucra ningún servicio proveído por el dueño del sitio web, al contrario, los dueños de los sitios buscan estrategias para **impedir** su uso.

Acceso **no controlado** se extrae información de forma anónima, por lo tanto, no se puede dar trazabilidad a las acciones de los usuarios

Requiere mayor limpieza de datos



API

Application program interface

Oficial

Mecanismo para extraer información de un sitio o aplicación web por medio de **protocolos de comunicación SOAP/REST.**

Servicio proveído por el **dueño** del sitio web.

Acceso controlado por el **dueño** del sitio, por medio de TOKENS o llaves de seguridad, con esto los identifica y puede dar trazabilidad a las acciones de los usuarios, imponer limitaciones y un mayor control.

Puede implicar un costo.

El objetivo de ambos es extraer información de sitios web

Limpieza de datos

Jornada
2022



2022/09/22 12:08

- Funciones de simplificación
- Diccionarios de conversión
- Expresiones Regulares
- Almsacnamiento en bases de datos (Sqlalchemy)

```
def formato_fecha(self,titulo_fecha):  
    ### Fecha formato###  
    from datetime import date  
    from dataclasses import replace  
    import re  
  
    if len(titulo_fecha)>4:  
        fullstring = titulo_fecha  
        substring = ":"  
        if substring in fullstring:  
  
            print(titulo_fecha)  
            fecha=titulo_fecha.replace(":", "").split()  
            del fecha[0]  
            del fecha[0]  
            mes= fecha[1]  
            dia= fecha[0]  
            año= fecha[2]  
            diccionario = {'enero':'01', 'febrero':'02', 'marzo':'03', 'abril':'04', 'mayo':'05',  
                            'junio':'06', 'julio':'07', 'agosto':'08', 'setiembre':'09', 'septiembre':'09', 'octubre':'10',  
                            'noviembre':'11', 'diciembre':'12'}  
  
            def Translate(resultado):  
                def SubTrans(mo):  
                    word = mo.group()  
                    return resultado.get(word.lower(), word)  
                return SubTrans  
  
            mes = re.sub(r'\w+', Translate(diccionario), mes)  
            print('el mes es',mes)  
            my_string=f"{mes}/{dia}/{año}"  
            print(my_string)  
            titulo_fecha=datetime.strptime(datetime.strptime(my_string, '%m/%d/%Y'), '%Y-%m-%d')
```

Almacenamiento

```
1 /***** Script for SelectTopNRows command from SSMS *****/
2 SELECT TOP (1000) [row_num]
3     ,[post_id]
4     ,[texto]
5     ,[cant_comentarios]
6     ,[cant_likes]
7     ,[like]
8     ,[angry]
9     ,[haha]
10    ,[care]
11    ,[love]
12    ,[sad]
13    ,[wow]
14    ,[reaction_count]
```

100 %

Results Messages

row_num	post_id	texto	cant_comentarios	cant_likes	like	angry	haha	care	love	sad
1	7534863926555909	Ciente del Banco Nacional que sufrió deducciones d...	1600	1030	1030	0	21	2	5	0
2	5796969470313747	El Banco Nacional anunció esta semana que las per...	25	59	59	4	13	1	7	1
3	5888862744474528	#Nacionales Aseguran que hay personas con men...	110	99	99	40	143	2	0	0
4	10159083372355334	¿Cuántos segmentos de usuarios de redes sociales ...	0	1	0	0	0	0	0	0
5	479420434028073	En honor a la verdad El suelto (opinión) colocado ay...	50	118	118	0	0	0	16	0
6	553768863424580	San José, 2 Nov (Elpaís.cr) - Los diputados de la Co...	0	0	0	0	0	0	0	0
7	3560018940891980	Una persona herida en vuelco esta noche en San Jo...	9	61	61	0	0	1	0	12



Casos y técnicas interpretación/análisis de datos

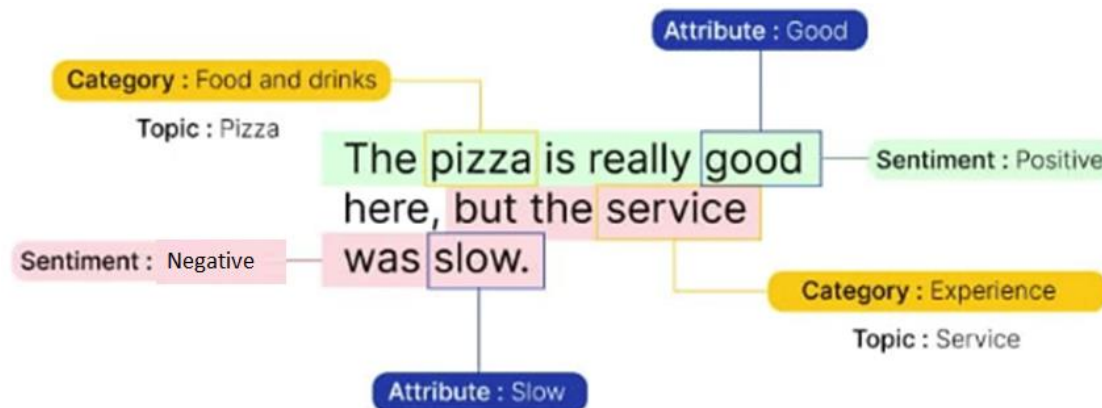
Jornada
2022

- Análisis de sentimientos
- Identificación de entidades
- Modelos personalizados de clasificación

Análisis de Sentimientos

Definición

- Se refiere al uso de procesamiento de lenguaje natural, análisis de texto y lingüística computacional para identificar y extraer información subjetiva de los recursos, evaluando las emociones, actitudes y opiniones.
- En términos más simples, clasifica un texto como positivo, negativo o neutral



Ventajas	Retos
Extracción de datos de forma automatizada	Estructura de datos
Disminuir cargas de trabajo	Limpieza y ordenamiento de datos
Manejo de grandes cantidades de datos	
Aumento en la velocidad de procesos	



¿Qué puede hacer el MOPT con sus obras clave si Meco y H Solís caen en insolvencia?

Las empresas constructoras Meco y H Solís, investigadas en el caso 'Cochinilla', tienen en su cartera a varios de los principales proyectos del Estado

```
from azure.ai.textanalytics import TextAnalyticsClient
from azure.core.credentials import AzureKeyCredential

# Authenticate the client using your key and endpoint
def authenticate_client():
    ta_credential = AzureKeyCredential(key)
    text_analytics_client = TextAnalyticsClient(
        endpoint=endpoint,
        credential=ta_credential)
    return text_analytics_client

client = authenticate_client()

# Example function for detecting sentiment in text
def sentiment_analysis_example(client):
    #print('esto es el texto', texto_cuerpo)
    texto_cuerpo1= texto[0:5000]
    documents = [texto_cuerpo1]
    response = client.analyze_sentiment(documents=documents)[0]
    print("Document Sentiment: {}".format(response.sentiment))
    print("Overall scores: positive={0:.2f}; neutral={1:.2f}; negative={2:.2f} \n".format(
        response.confidence_scores.positive,
        response.confidence_scores.neutral,
        response.confidence_scores.negative,
    ))

sentimientos= response.sentiment
```

Text Analytics

API ML Azure

Lenguaje:Python

La función de análisis de sentimientos proporciona etiquetas de sentimiento (como "negativo", "neutral" y "positivo") basadas en la puntuación de confianza más alta encontrada por el servicio a nivel de oración y documento

[Language Studio: Sentiment and opinion mining tryout \(azure.com\)](#)

id_sentimiento	url_noticia	sentimientos_azure	porcentajes_azure
3234	https://www.diarioextra.com/Noticia/detalle/484806/e...	neutral	Positivo: 0.1, Neutro: 0.84, Negativo: 0.05
3118	https://www.elfinancierocr.com/economia-y-politica/q...	negative	Positivo: 0.03, Neutro: 0.16, Negativo: 0.8
3235	https://www.larepublica.net/noticia/banco-nacional-no...	positive	Positivo: 0.95, Neutro: 0.04, Negativo: 0.01

Identificación de entidades

[link](#)

nacionales/fabrizio-alvarado-esteja-nombramiento-de-rector-de-la-uaca-en-el-cse/



Buscar



Últimas

Más Leídas

Nacionales

Deportes

Entretenimiento

Economía

Tecnología

Mundo

COVID-19

El nombramiento del rector de la UACA en el CSE ha dado de qué hablar, dado que el **órgano no sesionaba desde el pasado 27 de julio** porque no había quórum, ya que a Manuel Bolaños se le venció su designación.

“Qué bueno que al fin en el CSE tengamos **personas con criterios serios y sin sesgos ideológicos**, que aporten soluciones al apagón educativo que estamos viviendo. Qué bueno que se empiece a encender una luz en la educación”, concluyó el diputado.

Malavassi, al igual que Bolaños Salas, **fue jerarca del Ministerio de Educación Pública (MEP)**. Su administración se llevó a cabo desde **1966 hasta 1969**, durante el gobierno del expresidente José Joaquín Trejos.



Identificación de entidades

```

### Librerías para Spacy NER ###
import spacy
from spacy import displacy
from collections import Counter
import pandas as pd
pd.options.display.max_rows = 600
pd.options.display.max_colwidth = 400
nlp = spacy.load("es_core_news_sm")
doc = nlp(text)

for named_entity in doc.ents:
    if named_entity.label_ == "PER":
        people= named_entity.text+', '+people
        print(named_entity, named_entity.label_)

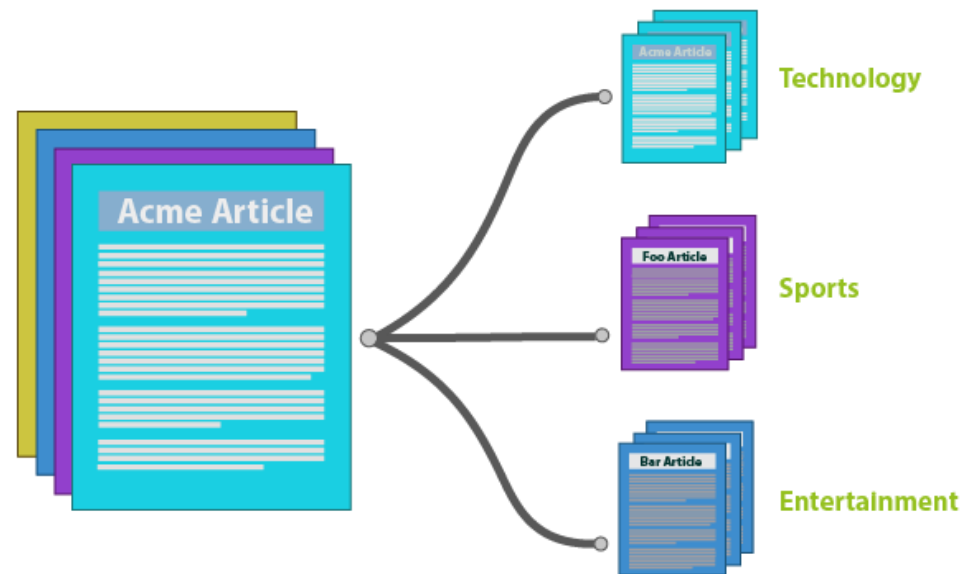
```

personas	localización	organización
Alvarado,Malavassi,Gobierno,Bolaños Salas,Manuel ...	Malavassi,La Gaceta,San José de Nueva República,...	Consejo,UACA,CSE,Consejo Nacional de Ed
Ricardo José Jiménez Paniagua,Rogelio Ugalde Delg...	Naranjo,También,BNCR,Sitecna,Vargas,Administraci...	Unidad de Administración del Espacio Aéreo,

Modelos personalizados de clasificación de texto

La clasificación de texto personalizado permite a los usuarios crear **modelos de IA para clasificar texto** en **categorías** predefinidas por el usuario.

El proceso busca etiquetar datos de forma iterativa, entrenar, evaluar y mejorar el rendimiento del modelo y disponerlo para su consumo.



Custom text classification

Train a classification model to classify text using your own data.

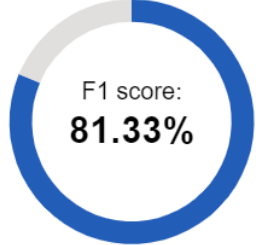
[Open Custom text classification](#)

<https://language.cognitive.azure.com/>

1. Creación y entrenamiento del modelo

2. Calibración

Status: ✔ Trained successfully



F1 score: 81.33%

Precision: 81.33%

Recall: 81.33%

Finished training on: 9/7/2022, 12:20:04 PM
 Total training time: 0 hour(s), 21 minute(s), 52 second(s)

Training data splitting type: Percentage
 Number of training documents: 664 (80%)
 Number of testing documents: 166 (20%)

[Learn more about how to improve your model](#)

3. Validación del modelo y clasificación de textos

Consultas

Confidence: 100.00%

Original text

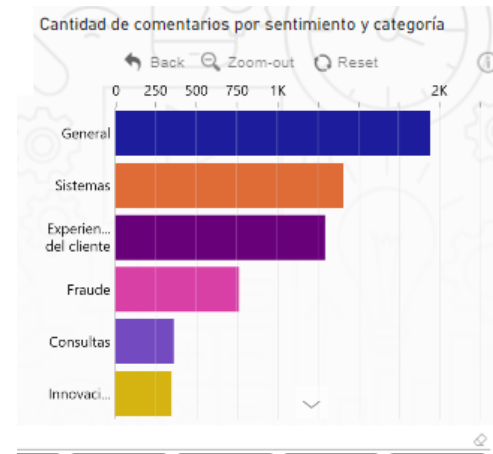
Cual es la taza de intrs de prstamos de vivienda?

Sostenibilidad

Confidence: 100.00%

Original text

Me preocupa la contaminación que se está generando



4. Medir los objetivos, analizar los resultados

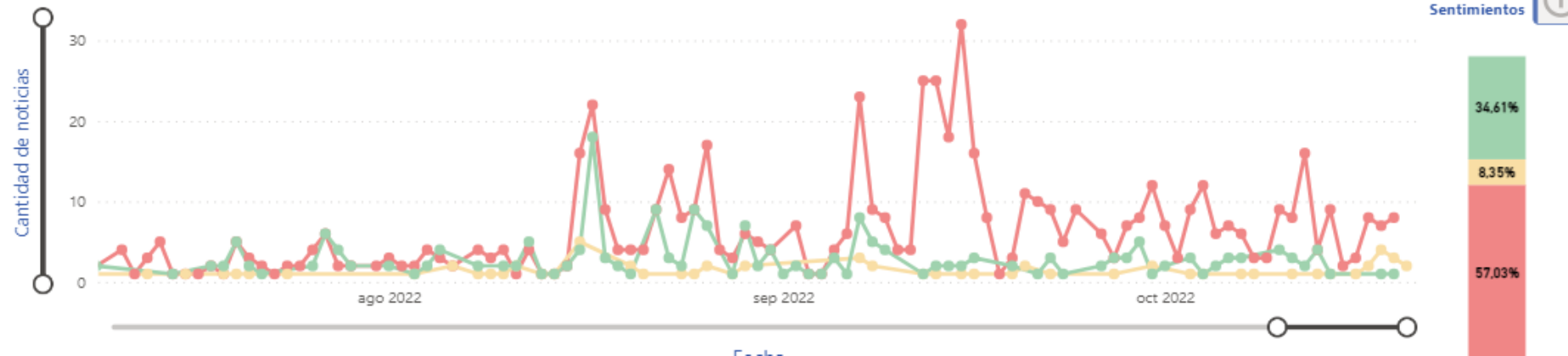
Escucha de Clientes				
Dimensiones				
Inteligencia Artificial Azure				
♥♥ Lovers ♥♥	Seguidores	xx Haters xx		
Felicitación	Comentario	Quejas		
Percepción positiva	Neutros	Percepción Negativa		
Modelo de Aprendizaje Supervizado				
Categorías	Rep. & Imagen	Experiencia del cliente		Op. & Tec. & Rep. Riesgos
		Sostenibilidad		
		Innovación		
		Seguridad		
		Fraude		
		Sistemas		
		Consultas		
		General		
		Referencias		
		Gen.		



Nota: Ver ejemplo desde Python

Interpretación de datos

Análisis de sentimientos de las notas



¿Qué nos dicen las noticias?



KPI Share of Voice

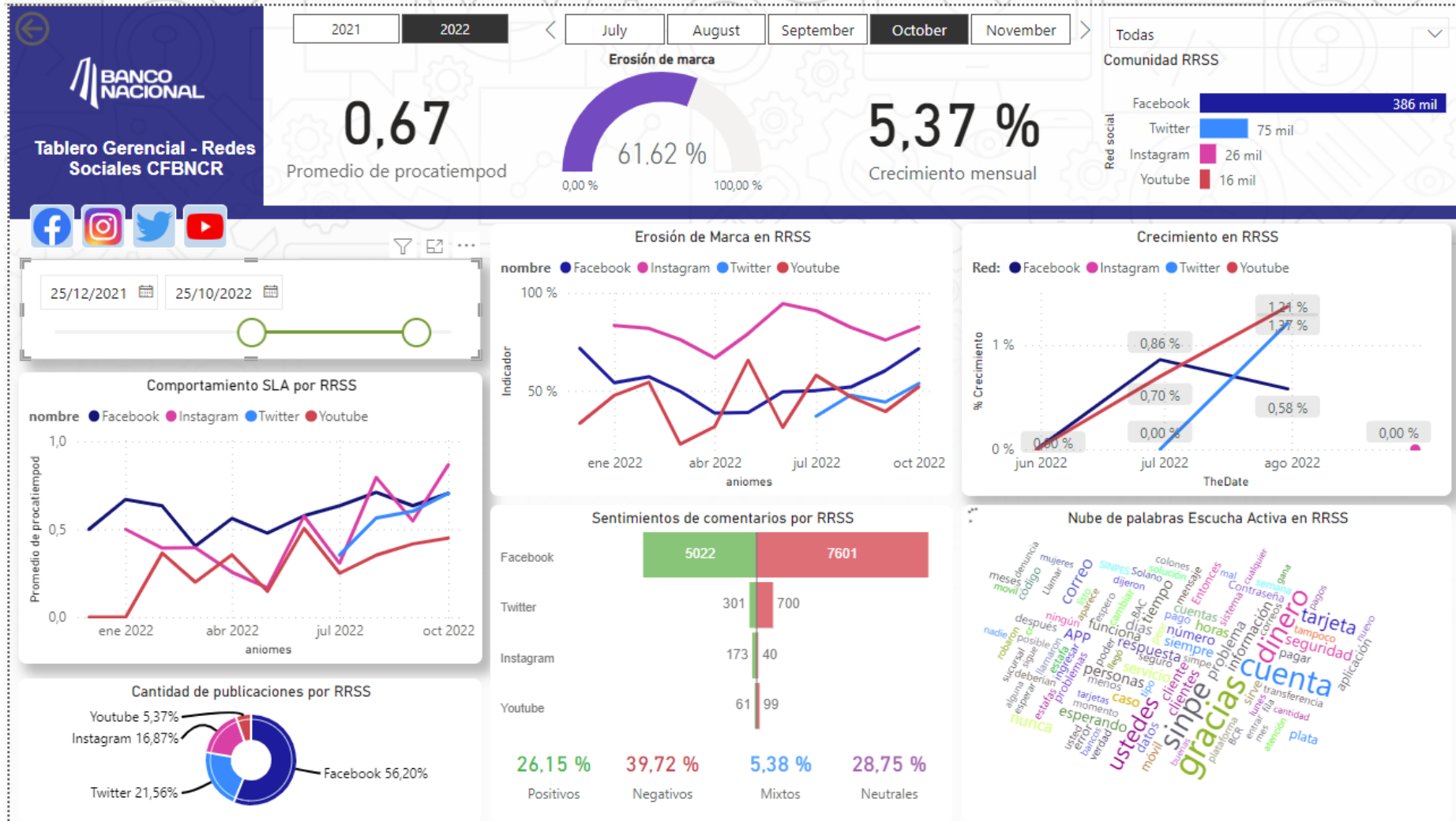
Noticias sobre el CFBNCR: 544 ➔ 24,30 %

544	60	45	61	89
BN_BN	BN_BNFondos	BN_BNSeguros	BN Valores	BN Vital

Noticias sobre la Competencia: 1824 ➔ 81,46 %

VF-Impacto de la Marca CFBNCR en los Medios - Power BI

Interpretación de datos



Ejemplo

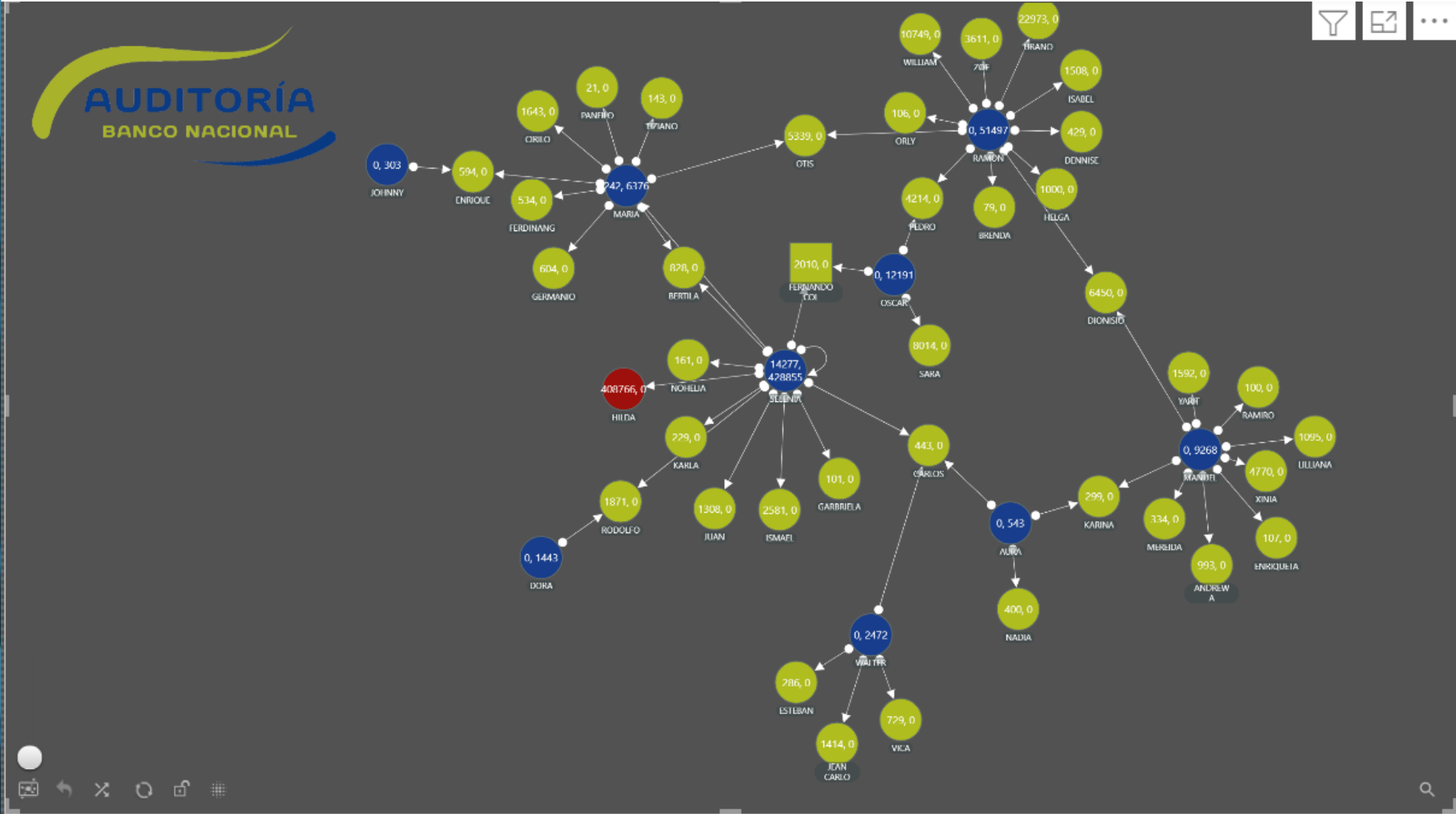
Noticia en la Nación Digital del 22 de julio 2022:

“Es una figura icónica en el crimen organizado y ha regentado durante muchos años el sector de María Reina de los Ángeles, en Hatillo, San José”. De esa forma describió Wálter Espinoza Espinoza, director del Organismo de Investigación Judicial (OIJ), a **un hombre (395)** de apellidos **Rodríguez Chaves (804)**, conocido como Pioja, un líder narco que fue detenido esta semana.

Aunque nunca se le ha conocido una actividad distinta al tráfico de drogas, Pioja, **de 49 años (13)** en la actualidad, logró abrirse paso en el comercio para legitimar el dinero que obtenía mediante la venta de drogas. Por ello, abrió una tienda en Hatillo, cuenta con casas de alquiler y poseía otros negocios.

También, a pesar de su bajo perfil, vivía en un **lujoso residencial en Sabana sur (1)** y no se limitaba a la hora de satisfacer a sus familiares con vehículos de lujo, joyas y dinero.

Ratcliff + Levenshtein = thefuzz



III. Lista de verificación

Jornada
2022

 **ISACA.**
Costa Rica Chapter

 Instituto de
Auditores Internos
de Costa Rica

Elementos a considerar

Jornada
2022

Procesos



Establezca el proyecto en función del riesgo, impacto

Procesos



Selecciona metodologías ágiles

Personas



Define equipo de trabajo multidisciplinario y colaborativo

Procesos



Definir los actores involucrados y sus expectativas

Tecnología



Analiza el entorno de los datos, accesos, permisos

Elementos a considerar

Jornada
2022

Personas



Establece el alcance del proyecto

Personas



Valorar contratación de un especialista o Partner

Tecnología



Contar con el software y/o hardware sea institucional o por parte de la Auditoría

Procesos



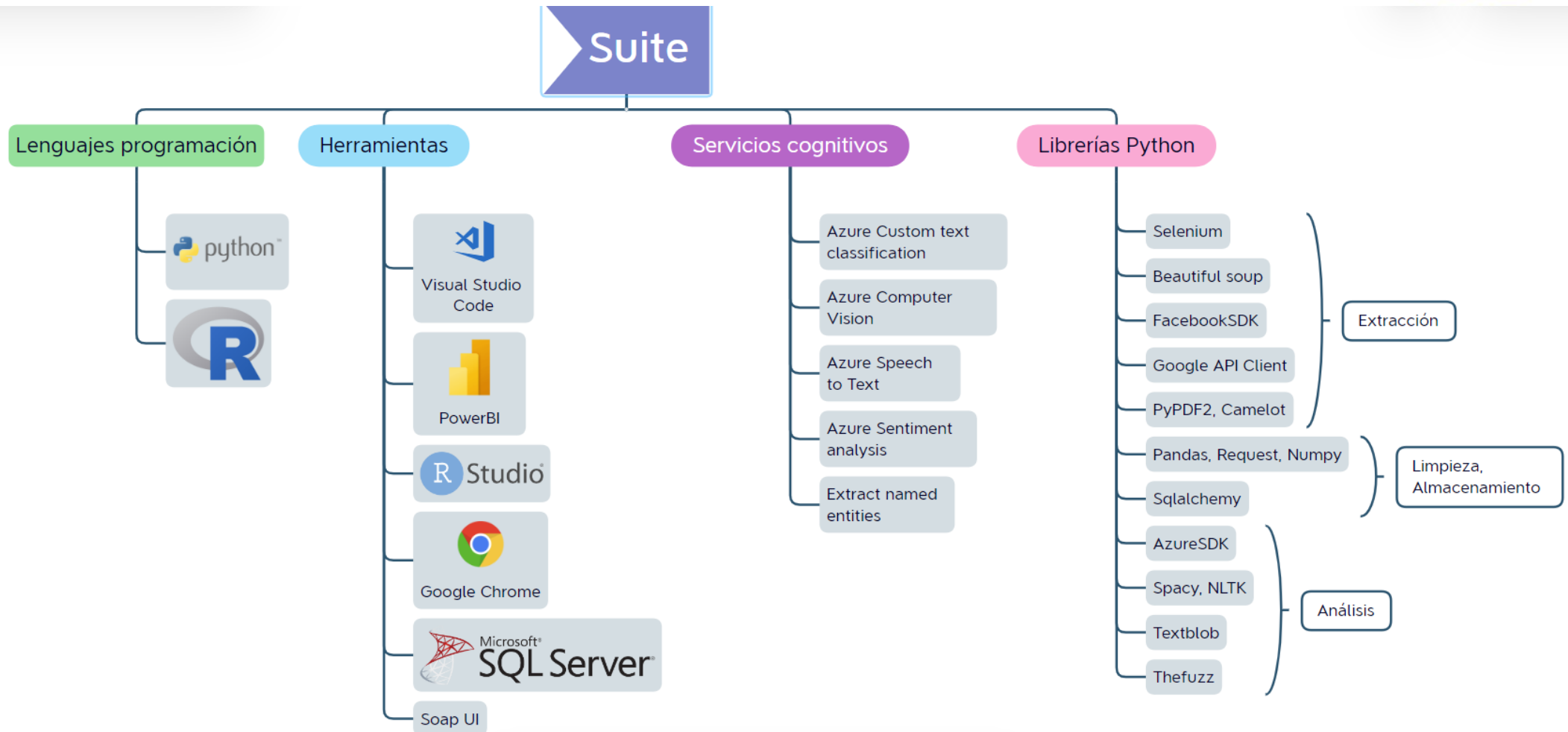
Definir Métricas

Procesos



Entrega de Producto y traslado de conocimiento

Recursos utilizados



Gracias

Jornada
2022

 **ISACA.**
Costa Rica Chapter

 Instituto de
Auditores Internos
de Costa Rica