

Prueba técnica

Vacante: Técnico de enlace

Dirección de Analítica y Ciencia de datos

Instrucciones: esta prueba consiste de cinco retos de programación y una pregunta de reflexión, por favor, atiende cada uno de los retos de acuerdo a tus habilidades y conocimientos.

Forma de entrega: debes crear un repositorio privado en GitHub y subir **un script por cada reto**. El repositorio lo debes compartir a los correos: bhvazquez@cjf.gob.mx y mahernandezroma@cjf.gob.mx.

Criterios a evaluar: debes describir de manera **clara y concisa** los pasos realizados para dar solución a cada uno de las retos. En el caso de la pregunta de reflexión, se sugiere cargar al repositorio un documento (PDF) con la reflexión realizada. Cada reto de programación se evaluará de acuerdo a los siguientes criterios:

- Procesamiento y análisis de datos (2 puntos)
- Diseño, implementación y ejecución del algoritmo (4 puntos)
- Análisis de resultados (2 puntos)
- Documentación de la solución (2 puntos)

Retos de programación

1. Construye un modelo de identificación de entidades nombradas usando las bibliotecas que consideres más apropiadas. Para el desarrollo de este reto debes usar datos abiertos gubernamentales. A continuación encontrarás algunas fuentes de datos recomendadas.

- [Tesis](#)
- [Tratados](#)
- [Comunicados](#)

En tu script debes indicar claramente la fuente de datos que estás utilizando.

2. Construye un modelo básico de preguntas y respuestas (desde cero, no se aceptarán modelos pre-entrenados). En el código deben ser claras las fases de preprocesamiento, entrenamiento, validación y pruebas. La selección de los datos de entrada es libre. En tu script debes indicar claramente la fuente de datos que estás utilizando.
3. Construir un API que coloque a disposición la información contenida en el archivo CSV [\[enlace\]](#). La información del archivo .csv deberá estar disponible a través de los siguientes endpoints:

- a) **GET** /ejercicio/descarga/
- b) **POST** /ejercicio/descarga/

Existirá un campo obligatorio llamado **formato** en el payload, mediante el cual, un usuario indicará el formato respuesta deseado, ejemplo:

```
POST /ejercicio/descarga/
{
  "formato": "csv"
}
```

GET /ejercicio/descarga/?formato=json

Dicho campo podrá aceptar alguno de los siguientes valores:

- **csv**: Responderá el archivo en formato csv
 - **json**: Responderá la información del en formato json.
4. Se han determinado las siguientes tuplas, las cuales representan una relación entre elementos jurídicos:

```
relaciones = [  
    ('Principio de Contradicción', 'Sentencia 338/2018'),  
    ('Juez de Amparo', 'Tribunal Colegiado'),  
    ('Auto de Vinculación a Proceso', 'Juez de Control'),  
    ('Organo Jurisdiccional', 'Juez de Control'),  
    ('Juzgado de Control', 'Sentencia 338/2018'),  
    ('Juzgado de Distrito', 'Sentencia 338/2018'),  
    ('Norma Convencional que contempla justa indemnizacion',  
    'Sentencia 423/2019'),  
    ('Autoridad Jurisdiccional', 'Suprema Corte de Justicia de la Nación'),  
    ('Organo Jurisdiccional', 'Juez de Amparo'),  
    ('Norma Convencional que contempla justa indemnizacion',  
    'Artículo 63.1 Convención Americana sobre Derechos Humanos'),  
    ('Juez de Amparo', 'Juzgado de Distrito'),  
    ('Tribunal Colegiado', 'Sentencia 338/2018'),  
    ('Norma Convencional que contempla justa indemnizacion',  
    'Norma Convencional'),  
    ('Prisión Preventiva Oficiosa', 'Órgano Jurisdiccional'),  
    ('Autoridad Jurisdiccional', 'Órgano Jurisdiccional')  
]
```

Realiza una función que reciba como entrada dos de los elementos jurídicos anteriormente mencionados (elementoJuridicoInicial, elementoJuridicoFinal) y como salida de una lista de elementos que muestre la ruta de conexión de los dos elementos dados. Ejemplo:

```
input(elemento1, elemento5) --> output: [elemento1, elemento3, elemento5]
```

5. Construye un sistema de recomendación usando PySpark. Para la lógica del sistema puedes emplear métodos de factorización de matrices o algún otro algoritmo que consideres pertinente. Altamente recomendable usar datos abiertos gubernamentales. En tu script debes indicar claramente la fuente de datos que estás utilizando.

Reto de reflexión

1. Actualmente el gobierno está haciendo un esfuerzo para contribuir en la transformación digital del país. Con tu experiencia y habilidades actuales, ¿qué soluciones basadas en ciencia de datos propones que sean posibles de implementar y que beneficien a la sociedad mexicana?.