

Prueba práctica

En el archivo adjunto (zip) se encuentran los textos extraídos de algunos planes de desarrollo territorial – PDT (documentos que trazan la política pública de los municipios y departamentos).

La lectura no es exacta debido a que estos son reportados en formato pdf, algunos son documentos escaneados, la presencia de tablas y esquemas por lo cual no existe exactitud con el texto original.

Sin embargo, con esta información es posible utilizar métodos de procesamiento natural de lenguaje que permitan caracterizar los planes de desarrollo territorial. Esto con el objetivo de brindar información a las personas encargadas de evaluar los PDT.

El propósito de esta prueba es diseñar e implementar métodos que permitan evaluar los planes bajo los siguientes aspectos:

1. Análisis de temas principales en cada PDT
2. Clasificar los planes de acuerdo con algún criterio de similitud
3. ¿Qué tipo de información reportaría a los evaluadores con el fin de hacer su tarea más sencilla?

Se debe enviar un informe que contenga la siguiente información:

1. Descripción de cada uno de los pasos para el procesamiento del texto.
2. Modelos de clasificación utilizados.
3. Criterio de selección para determinar cuál sería el mejor clasificador.
4. Validación de datos utilizada.

Se puede hacer uso de lenguaje R o Python. El script resultante (que contenga todos los procedimientos) debe ser enviado junto con el informe.