

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



ANÁLISIS EXPLORATORIO Y DESCRIPTIVO DE LOS PDT COSTEROS Y ALINEAMIENTO CON EL CONPES 3990

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital - Unidad de Científicos de Datos
- Dirección de Descentralización y Desarrollo Regional

Otras entidades

Sector

Inversión y finanzas públicas

Lenguaje

Python

Fuente de datos

Planes distritales territoriales (municipios costeros) y CONPES 3990

Presentación

La Dirección de Descentralización y Desarrollo Regional (DDDR) ha planteado la necesidad de realizar un análisis a los Planes Distritales Territoriales (PDT) de 46 municipios costeros para identificar problemas, ejes estratégicos y programas planteados en los PDT. Para esto se planteó realizar un análisis de texto a los PDT de acuerdo con divisiones específicas a cada documento con base en las indicaciones proporcionadas por la DDDR, adicionalmente realizar un análisis comparativo de cada PDT con respecto al CONPES 3990 “Colombia potencia bioceánica sostenible 2030” para identificar que tan similares son los PDT con este documento.

The DJSG presents a proposal for a justice viewer that arises from its effort in the consolidation of justice information from various sources such as DANE, DNP, entity files and Google. In such a way that the UCD developed a viewer with said information with the PowerBI platform and under the parameters and criteria established by the DJSG for its publication to the general public. With this viewer, it is expected to have a consolidated and practical system for consulting judicial entities in Colombia and to have basic descriptive statistics that allow giving a summarized view of the state of judicial entities in Colombia.

Objetivo general

Realizar una revisión de temas generales en los Planes Distritales Territoriales (PDT) del periodo 2020-2024 correspondientes a 46 municipios costeros para identificar problemas, ejes estratégicos y programas planteados en los PDT, de acuerdo con las indicaciones establecidas por la DDDR.

Objetivos específicos

1. Realizar análisis de texto a los PDT de acuerdo con divisiones específicas a cada documento con base en las indicaciones de la DDDR
2. Realizar un análisis comparativo con cada documento PDT con respecto al CONPES 3990 “Colombia potencia bioceánica sostenible 2030” para identificar qué tan similares son los PDT con este documento.

Metodología

Para el desarrollo de este proyecto, se abordó el problema por medio de diferentes pasos



1. Lectura de todos los PDT

Usando la librería de Contexto desarrollada por la UCD con métodos de lectura y escritura para trabajar con textos limpios y así posteriormente poder hacerle limpieza a los PDT y eliminar contenido que puede resultar afectando al momento de trabajar con esos textos limpios los procesos siguientes.

2. Identificación de Stopwords

Un tipo de stopwords son una serie de palabras que son comunes dentro del idioma que (entre otras cosas) pueden servir para darle un sentido gramatical a las oraciones. Estas palabras se caracterizan por darle un sentido a la oración. Sin embargo, para este análisis, no aportan información. Un ejemplo de ello son conectores, artículos, preposiciones, etc., todas palabreas que tienen importancia dentro de la oración dentro de su sintaxis, pero no hacen parte de las palabras clave de estas. Adicional a esto, estas palabras son comunes en las oraciones por lo que, al realizar un análisis de palabras, estas palabras van a tener una importancia alta dado que su ocurrencia es transversal a todas las oraciones.

Primero se utilizaron listas predefinidas de stopwords para el idioma español que se encuentran en internet. Sin embargo, existen palabras muy específicas que pueden considerarse como stopword dado que es común pero no aporta más al análisis, estas últimas dependen del objetivo del análisis y se van a tomar basados en ello. Un ejemplo de estas palabras son nombres de países, departamentos, ciudades, nombres de páginas web y en general palabras que no aportan al análisis.

Para encontrar estas palabras se hace por medio de las frecuencias de las palabras, las palabras con mayor frecuencia suelen ser este tipo de palabras, sin embargo, esta selección se hace con intervención humana, dado que, si se eliminan las n palabras con mayor frecuencia, podría estar eliminando una gran cantidad de stopwords pero también se podrían estar eliminando palabras que puedan aportar al análisis. Esta lista de palabras se adquiere una vez y se guardan en una lista para ser utilizadas cada vez que se necesite.

Una vez se identificaron las stopwords, se realizó una limpieza de texto en las cuales incluye la eliminación de estas palabras, estandarización de palabras, eliminación de caracteres especiales, entre otros. Esto con el fin de estandarizar el texto y facilitar la implementación de los modelos.

3. Visualizaciones

Con base el texto limpio, se realizaron nubes de palabras, gráfico de frecuencia y gráfico de co-ocurrencia per permite dar una visión general del contenido del teto. Esto se realizó individual en cada sección del texto (diagnóstico, estratégico e inversión) de tal modo que hay una visualización individual por cada uno de los rubros mencionados. Esto con el fin de identificar las palabras predominantes dentro de cada sección y tener una visión general del contenido y temas del texto.

4. Creación del reporte

Se realizó un reporte por cada PDT analizado, el cual contiene los resultados organizados. De igual manera, se enuentra dividido por sección dentro del PDT (diagnóstico, estratégico e inversión). Este reporte se realizó con un template base y se organizaron los resultados con la herramienta Jinja de Python. Esto permite crear un reporte individual en HTML por PDT. Por su parte, cada reporte tiene enlazados los otros reportes, lo cual facilita la navegación entre reportes de acuerdo al PDT de interés.



Resultados

Posterior a la metodología descrita en la sección 3 se obtuvieron los resultados que se presentan a continuación. Todos estos resultados se consolidaron en un HTML que permite a la persona filtrar el departamento y el municipio del cual desea visualizar los resultados. Adicionalmente se le agregó un nivel de similitud, esto en comparación con el CONPES 3990 y el PDT que se está visualizando, tal y como fue solicitado por la DDDR.

El HTML presenta la información de cada uno de los municipios como se muestra en la Ilustración 1, 2, 3 y 4. En la Ilustración 1 podemos observar la vista general del aplicativo, con las herramientas de filtrado, nivel de similitud y la estructura final del reporte del municipio elegido en el lado derecho de la ilustración.

Ilustración 1 Vista General Aplicativo



Fuente: elaboración propia

En la *Ilustración 2* se puede observar el filtro para este aplicativo. Principalmente se elige el departamento del cual queremos elegir posteriormente su municipio, automáticamente se presentará la información del municipio que se eligió, donde aparecerán sus respectivas nubes de palabras, gráficos de frecuencias, y gráficos de co-ocurrencia, además de su nivel de similitud. Como se presenta en la *Ilustración 2* después de la selección del municipio podemos observar el nivel de similitud para nuestro municipio elegido. Este nivel de similitud nos permite ver una comparación entre el PDT municipal y el CONPES 3990.

Ilustración 2 Filtros aplicativo

Departamento

Antioquia

Municipio

Necolí

Nivel de similitud entre
PDT municipal y CONPES
3990

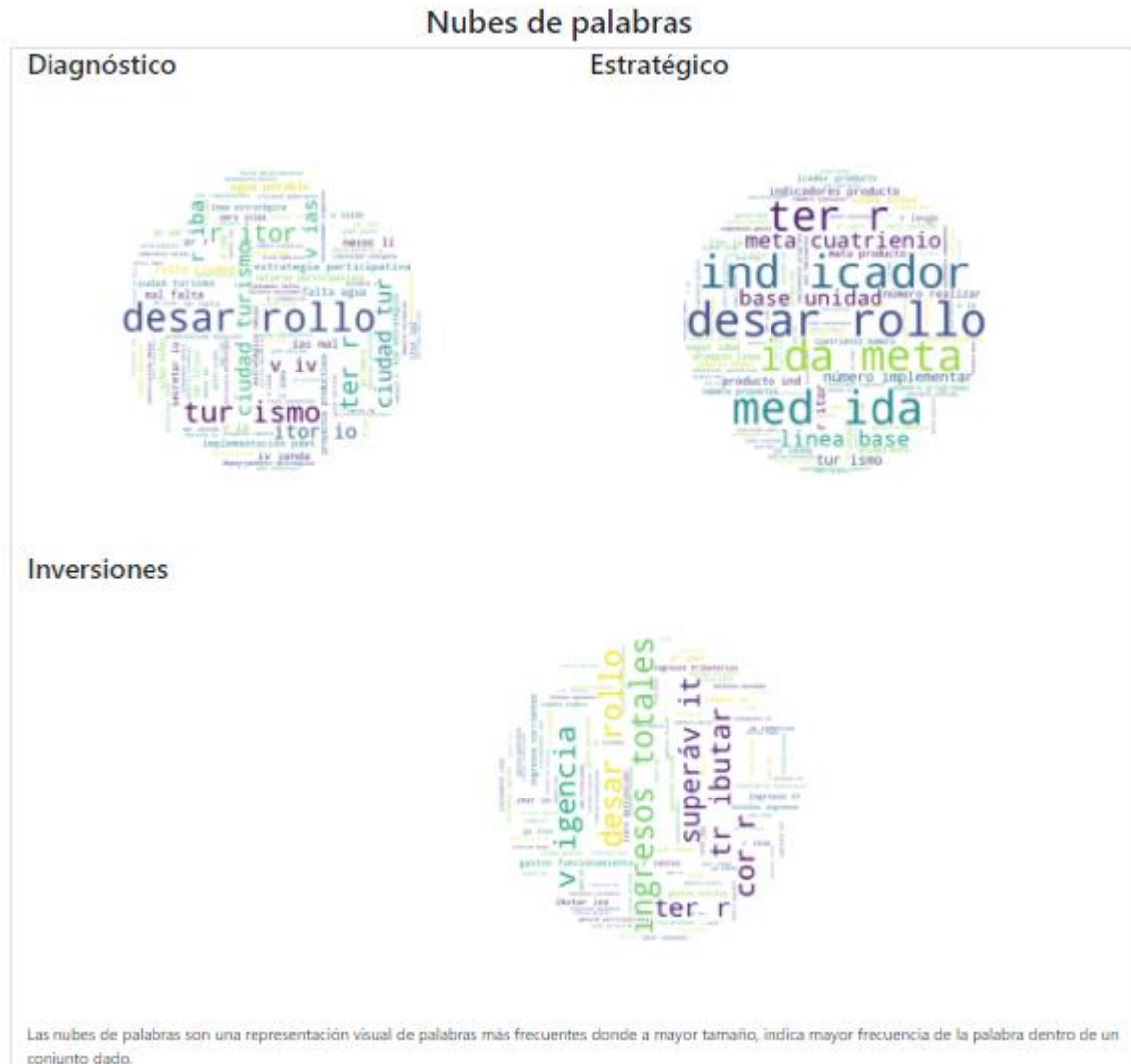
0.2467

Fuente: elaboración propia

En la *Ilustración 3* podemos observar la estructura de las nubes de palabras para su comparación, donde cada nube de palabra hace parte de la subdivisión solicitada por la DDDR, las cuales son Diagnóstico, Estratégico e Inversiones. Se planteó que las nubes de palabras estuviesen juntas para su respectiva comparación entre las diferentes subdivisiones.



Ilustración 3 Sección Nubes de palabras

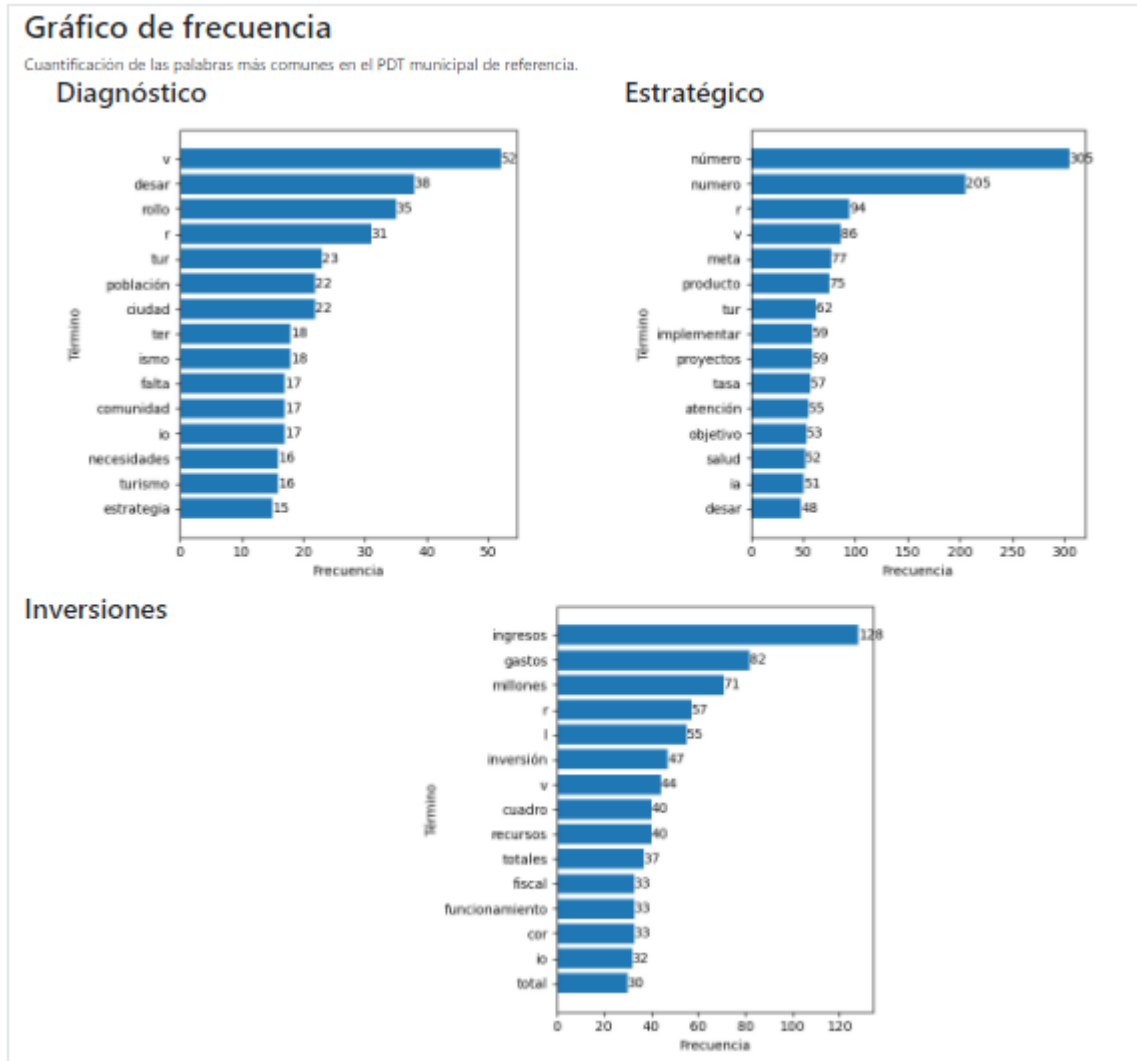


Fuente: elaboración propia

Para la *Ilustración 4* observamos la misma estructura que en la sección de nubes de palabras en la *Ilustración 3*, donde se ven las gráficas de frecuencia para cada una de las subdivisiones de Diagnóstico, Estratégico, Inversiones, lo que permite comparar por subsección del PDT municipal, de qué se habla en mayor medida. Los gráficos de frecuencias son gráficos que permiten identificar las palabras “clave” y así poder entender mejor de qué se habla en los PDT y en cada subsección. La lectura del gráfico se realiza observando desde arriba hacia abajo, los elementos que están más arriba son los que se usan con mayor frecuencia dentro del PDT, esto puede facilitar al momento de entender los puntos fuertes de la subdivisión.



Ilustración 4 Sección Gráfico de frecuencias

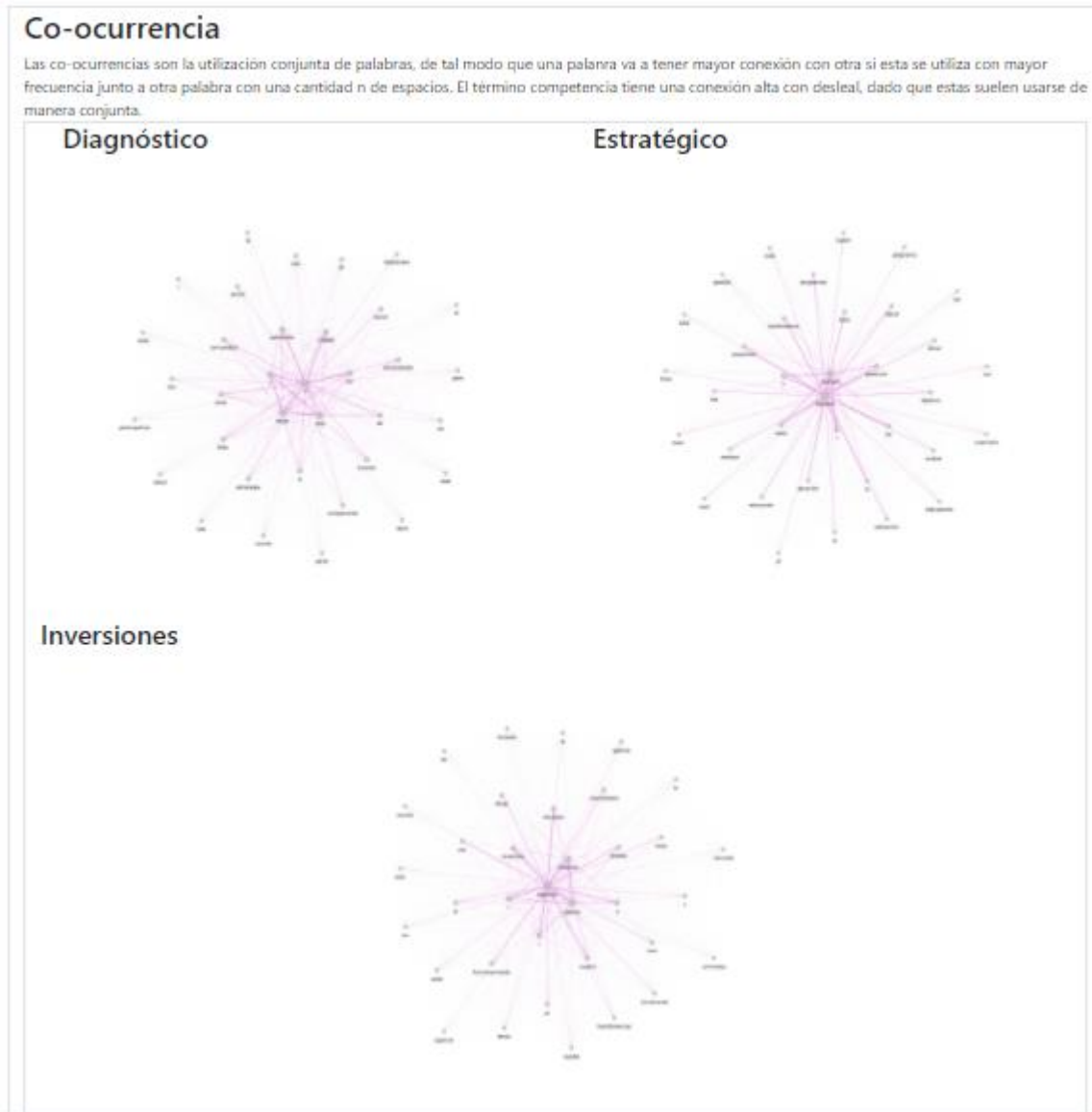


Fuente: elaboración propia

Finalmente está en la *Ilustración 5* tenemos la sección de co-ocurrencias, donde podremos observar cómo se conectan las palabras, con qué palabra es más frecuente su aparición, y qué tan frecuente es por el tamaño del punto. Este gráfico de co-ocurrencia nos permite identificar cuales palabras son más frecuentes que aparezcan en función de otra, por ejemplo para nuestra tenemos que en Diagnóstico la palabra “población” se conecta con “figura” lo que quiere decir que en muchos casos que aparezca población, aparecerá figura ya sea antes o después de población.



Ilustración 5 Gráficos de co-ocurrencia



Fuente: elaboración propia

A través del desarrollo metodológico descrito en la Sección **¡Error! No se encuentra el origen de la referencia.**, se obtuvieron los resultados que se presentan a continuación. Toda retroalimentación desde un punto de vista experto o de usuario por parte de la (nombre de la dependencia / entidad) es bienvenido. Este insumo será de gran ayuda para mejorar la calidad y utilidad de los resultados obtenidos, de manera que agreguen mayor valor.



Conclusiones y recomendaciones

1. Se desarrollaron reportes individuales por cada PDT con los resultados de cada sección encontrada en el mismo. Así mismo, los reportes tienen enlaces uno con el otro, lo cual facilita navegar entre los resultados de los PDT.
2. La comparación de texto con el PDT y el CONPES 3990 se realizó con diferentes técnicas, sin embargo, la decisión final de cual es la mejor técnica depende de la DDDR, siendo esta la final para mostrar en los reportes.
3. La información de cada reporte es estática, lo que quiere decir que no se actualiza y los resultados son únicos por PDT.

Socialización

Se ha socializado con la UCD de la DDD.