

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



ANÁLISIS DE LOS PLANES DE GOBIERNO DE LOS NUEVOS MANDATARIOS

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital.
- Dirección Descentralización y Desarrollo Rural.

Sector

Territorial

Lenguaje

Python, R y Markdown

Fuente de datos

Planes de gobierno, lista de alcaldes y gobernadores (DDDR), PND y página web de Naciones Unidas.

Presentación

Los planes de desarrollo territorial (PDT) son los instrumentos de planificación con los cuales las administraciones locales definen los programas y proyectos que ejecutarán durante su período de gobierno. Dada la importancia estratégica de los PDT, es de interés nacional que su formulación sea acorde con las necesidades de las entidades territoriales y con los lineamientos establecidos en la planeación nacional, para lo cual este proyecto propone una herramienta que permite identificar los temas de mayor relevancia en los PDT, estimar el nivel de alineación de los PDT con los ODS y estimar el nivel de alineación de los PDT con el Plan Nacional de Desarrollo (PND). La herramienta puede ser utilizada por alcaldes y gobernadores para orientar la realización de los PDT en sus correspondientes entidades territoriales.

Territorial development plans (PDTs, for its Spanish acronym) are the planning instruments with which local administrations define the programmes and projects they will implement during their period of government. Given the strategic importance of the PDTs, it is of national interest that their formulation be in accordance with the needs of the territorial entities and with the guidelines established in the national planning, reason why this project proposes a tool that makes it possible to identify the most relevant subjects in the PDTs, to estimate the level of alignment of the PDTs with the ODS and to estimate the level of alignment of the PDTs with the National Development Plan. This tool can be used by mayors and governors to guide the implementation of the PDTs in their corresponding territorial entities.

Objetivo general

Hacer más sencilla para alcaldes y gobernadores la formulación de sus Planes de Desarrollo Territorial, a través de una herramienta de visualización.

Objetivos específicos

1. Realizar la lectura automática de los planes de gobierno de aspirantes a alcaldías y gobernaciones.
2. Identificar las temáticas más importantes de los programas de gobierno mediante un análisis descriptivo de los textos.
3. Estimar el nivel de alineación del plan de gobierno con los Objetivos de Desarrollo Sostenible y con el Plan Nacional de Desarrollo a partir del cálculo de una medida de similitud entre textos.

Metodología

El desarrollo de este proyecto se dividió en tres fases: el análisis descriptivo, el cálculo de medidas de similitud y la consolidación de los resultados en una herramienta de visualización. El detalle de cada etapa se muestra a continuación.



Análisis descriptivo

Para el análisis descriptivo de los planes de gobierno se realizó la lectura automática de los PDF utilizando Python, con lo que se obtuvo una cadena de texto para el contenido de cada página. Algunos documentos requirieron del paso por un OCR (reconocimiento óptico de caracteres), un paso necesario pero que puede generar imprecisiones en la lectura. La limpieza de las cadenas de texto obtenidas se realizó con R y consistió en la transformación del texto a minúsculas y en la remoción de números, signos de puntuación y demás caracteres que no fueran letras; también se removieron conectores, preposiciones y palabras que no agregan significado al texto, entre las cuales se incluyeron el título del plan de gobierno, el nombre de la entidad territorial y el nombre del candidato (estas últimas se eliminaron únicamente del plan de gobierno correspondiente, evitando que se eliminaran palabras como “paz” en los planes de las entidades territoriales diferentes de “La Paz”). Para este proyecto no se realizó la remoción de tildes, pues la inclusión de palabras como “género” (que sin tilde se relacionaría con “generar”) era relevante en este análisis. Se resalta, sin embargo, que los acentos son una de las principales limitaciones que se tienen al utilizar un OCR para leer los documentos. El resultado de este tratamiento es lo que se denominará “texto limpio”.

Habiendo completado la limpieza del texto, se realizó un conteo de palabras (unigramas) y de pares de palabras (bigramas) para determinar cuáles aparecían con mayor frecuencia, resultado que se utilizó como indicador de los temas principales que se tratan en cada plan de gobierno. Adicionalmente, se realizó un análisis de coocurrencia de términos, que es un conteo de las veces que aparece cada posible par de términos en una misma página. Este conteo de coocurrencias permite establecer relaciones entre palabras y construir redes con las cuales un usuario puede, de manera sencilla, hacerse una idea de los temas principales a partir de la lectura ordenada de las palabras más frecuentes (tamaño del nodo en la red) y de las que tienen relaciones más fuertes (grosor del enlace en la red). La figura 1 presenta un ejemplo de una nube de bigramas y de una red de palabras, elementos utilizados para presentar los resultados del análisis descriptivo.

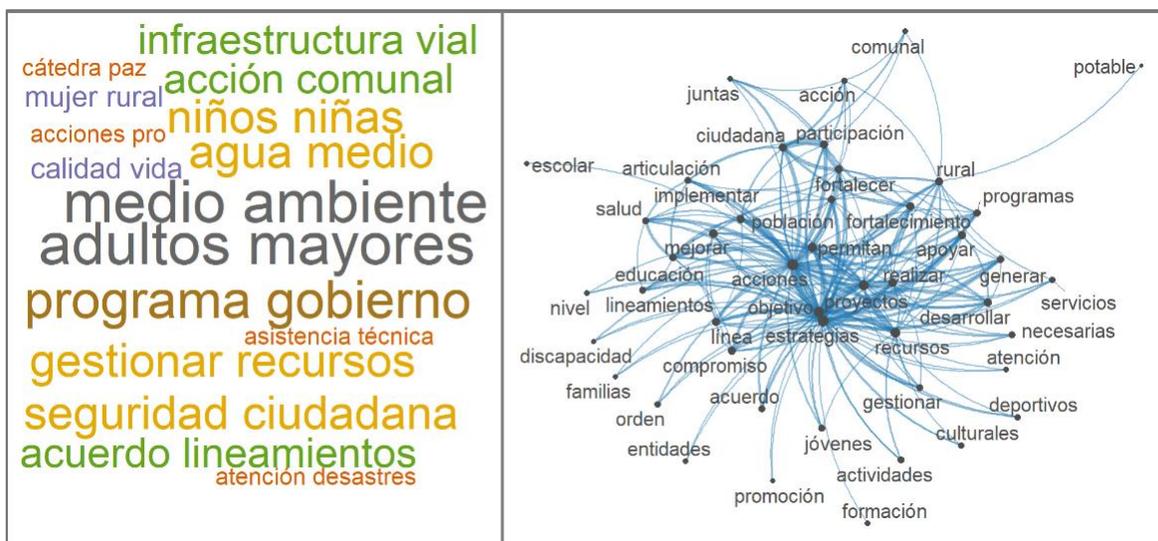


Figura 1: Nube de bigramas (izquierda) y red de palabras (derecha) construidas con base en un plan de gobierno.



Cálculo de medidas de similitud

Para la segunda fase del proyecto se comenzó consolidando, en cadenas únicas, los planes de gobierno, los textos disponibles de cada ODS y los pactos del PND. En el caso de los planes de gobierno y de los pactos del PND, bastó con unir los textos de todas las páginas correspondientes. En el caso de los ODS, en cambio, fue necesario juntar textos de diferentes tipos, que fueron (1) los textos descriptivos y datos destacables de cada ODS, disponibles públicamente en el enlace <https://www.un.org/sustainabledevelopment/es/objetivos-de-desarrollo-sostenible/>; (2) los documentos sobre “¿Por qué es importante?” relacionados con cada ODS y disponibles en el mismo enlace; (3) el detalle de objetivo, metas e indicadores, contenido en la base de datos que se ha utilizado como insumo en proyectos anteriores sobre ODS, y (4) las palabras clave que se identificaron en el proyecto “Rastreo de recursos para los ODS “ (cuyo informe se encuentra disponible en <https://colaboracion.dnp.gov.co/CDT/Desarrollo%20Digital/Big%20Data/02%20ODS/02%20Informe.pdf>). Para este análisis se excluyó el ODS 17 (alianzas para alcanzar los objetivos) ya que este es transversal a los demás objetivos e involucra relaciones en el ámbito nacional e internacional, lo que dificulta su caracterización e identificación.

Habiendo obtenido las cadenas de texto consolidadas, se realizó una vectorización (es decir, una representación numérica) de los textos de cada ODS. Para ello, se utilizó el modelo de bolsa de palabras (*bag of words*), que consiste en construir una matriz con palabras en las columnas y ODS en las filas que representa el número de veces que aparece cada palabra en el texto de cada ODS. Por ejemplo, si el texto del ODS 1 fuera únicamente “acabar pobreza extrema”, las columnas “acabar”, “pobreza” y “extrema” tendrían un valor de 1 en la primera fila de la matriz, mientras las demás columnas tendrían un valor de 0. Esta representación numérica se ilustra con mayor claridad en la figura 2. Cabe mencionar que esta técnica es bastante sencilla en comparación con otras como doc2vec y BERT, con las cuales podría tenerse una metodología más cercana al estado del arte, pero esta fue suficiente para satisfacer los requerimientos expresados por la dirección solicitante.

	este	es	un	texto	otro	último
este es un texto	1	1	1	1	0	0
este texto es otro texto	1	1	0	2	1	0
un último texto	0	0	1	1	0	1

Figura 2: Representación numérica de textos con el modelo de bolsa de palabras.

Habiendo realizado la vectorización de los textos de ODS, se procedió a vectorizar los planes de gobierno con base en el diccionario de ODS (palabras presentes en los textos de ODS), pues era esencial que los vectores finales fueran comparables, para lo que se requería que estuvieran en el mismo espacio vectorial. En términos prácticos, la realización de este procedimiento sería equivalente a eliminar de los planes de gobierno las palabras que no estuvieran en los textos de los ODS, para luego vectorizarlos fácilmente como nuevas filas en la misma matriz.

Posteriormente, para cada plan de gobierno se procedió a calcular la similitud coseno entre el vector “plan de gobierno” y cada uno de los vectores “ODS”, obteniendo 16 medidas de similitud para cada plan de gobierno (una con respecto a cada ODS). Esta medida de similitud se escogió porque ha sido ampliamente utilizada en análisis de texto y se calcula con un bajo costo computacional, criterio de gran importancia considerando que el algoritmo final debía ejecutarse para todos los municipios y departamentos de Colombia. Finalmente, se calculó el promedio de las 16 medidas para tener una medida agregada, lo que corresponde a un indicador de



alineación con todos los ODS que, aunque no tiene un significado matemático (como sí lo tienen las 16 medidas), tiene un gran valor a nivel práctico.

Para el cálculo de la similitud entre los planes de gobierno y los pactos del PND se realizó exactamente el mismo procedimiento que entre planes y ODS.

Consolidación de los resultados en una herramienta de visualización

Habiendo concluido el análisis, se diseñó una plantilla en HTML utilizando RMarkdown y flexdashboard, librerías de R que facilitan la integración de distintos elementos de texto y gráficos con los cuales el usuario puede interactuar desde su navegador. Para los gráficos de barras, con los cuales se ilustró la similitud con cada ODS y con cada pacto del PND, se utilizó la librería Plotly, lo que permite que el usuario conozca el nombre de cada ODS y de cada pacto al pasar el cursor sobre el gráfico.

Resultados

Se realizó el procesamiento de todos los planes de gobierno de alcaldes y gobernaciones, según la metodología indicada, y se generó un reporte para cada uno de ellos. Las figuras 3 y 4 ilustran el resultado final.



Figura 3: Vista del módulo de similitudes del reporte final.



El futuro es de todos

DNP
Departamento
Nacional de Planeación



Figura 4: Vista del módulo de palabras más frecuentes del reporte final.

Conclusiones

1. Se desarrolló una herramienta que orienta a los alcaldes y gobernadores en la formulación de sus planes de desarrollo territorial.
2. Los análisis descriptivos realizados facilitan la identificación de temas a los cuales se da relevancia en cada plan de gobierno.
3. El algoritmo desarrollado permite estimar el nivel de alineación de los planes de desarrollo con los ODS y con los pactos del PND.

Socialización

El proyecto se socializó con la Dirección de Descentralización y Desarrollo Rural.