

DIRECCIÓN DE DESARROLLO DIGITAL

Unidad de Científicos de Datos



DNP Departamento
Nacional
de Planeación



GOBIERNO DE COLOMBIA

Unidad de Científicos de Datos
ucd@dnp.gov.co

Metas PDT

Tabla de contenido

1. <i>Introducción</i>.....	2
2. <i>Objetivo</i>.....	2
3. <i>Descripción de los datos</i>	2
4. <i>Metodología</i>.....	2
5. <i>Resultados</i>	4
6. <i>Referencias</i>	6

1. Introducción

El Sistema de Información para la Evaluación de la Eficacia (SIEE) contiene los planes de desarrollo territorial (PDT), estos son instrumentos de planeación que orientan las acciones que se llevaran a cabo durante el gobierno en una entidad territorial, se determina la visión, programas, proyectos y metas asociadas a los recursos públicos.

Las metas de los PDT son redactadas por los gobernantes sin seguir lineamientos, esto dificulta comparar el progreso de las entidades territoriales de acuerdo con el progreso y cumplimiento de las metas, programas y proyectos planteados en los PDT. En este documento se encuentra una descripción detallada del análisis que se realizó sobre las metas de los PDT del 2016. Como complemento a este archivo hay un proyecto en R llamado "Terridata_Metas" que contiene los insumos, scripts y resultados del análisis.

2. Objetivo

Estandarizar las metas de los PDT, usando herramientas de análisis textual, con el fin de crear metas tipo por sector.

3. Descripción de los datos

La base de datos contiene 253.237 metas de los planes territoriales del año 2016, también se encuentra el departamento, el municipio y el sector al que pertenece cada meta. También se utilizarán las tipologías de cada entidad territorial para realizar análisis por sector y tipología. La base de datos de las metas está disponible en el SIEE.

4. Metodología

La metodología utilizada para el análisis de las metas consta de 5 pasos:

1. Preprocesamiento:

En este paso se remueven siglas, nombres geográficos y Stopwords, se remueven en total 1.106 siglas, 11.476 nombres geográficos y 901 Stopwords, remover los nombres geográficos es importante, ya que las metas siempre van dirigidas a un espacio geográfico y lo que pretende el proyecto es estandarizar las metas sin tener en cuenta el espacio geográfico. Las siglas que se remueven en este paso fueron validadas por la Dirección de Descentralización y Desarrollo Regional (DDDR), y también se añadieron unas excepciones de siglas. Las Stopwords son palabras vacías que carecen de significado en el contexto del procesamiento natural del lenguaje. Se muestran 10 palabras de cada uno de los grupos como ejemplo en las siguientes tablas.

Nombres geográficos
colombia
república de colombia
amazonas
antioquia
arauca
atlántico
bogotá
bolívar
boyacá
caldas

Stopwords
de
la
que
el
en
y
a
los
del
se

Siglas
IDECA
FAMIS
IDPAC
SIDAP
SIGOB
COTSA
ZODES
IFTDH
PRAUS
RCLPD

2. Revisión de palabras frecuentes:

Con el texto preprocesado se realiza un conteo de palabras y se hace una revisión de las palabras más frecuentes para determinar cuáles de estas en el contexto de las metas carecen de significado para el análisis que se desea realizar. Algunas de las palabras se muestran en la siguiente tabla:

Palabras frecuentes
implementar
programa
proyecto
plan
apoyar
población
público
estrategia
acción
proceso

3. Lematización:

Con el texto revisado se utiliza el lematizador de la RAE para llevar las palabras a su forma básica. Esto elimina plurales y coloca los verbos en infinitivo (terminación ar-er-ir). Se muestra un fragmento del lematizador de la RAE.

Original	Corrección
mantener	mantener
abandere	abanderar

abandonada	abandonar
abandonadas	abandonar
estrategia	estrategia
prevención	prevención
integral	integral
fortalecimiento	fortalecimiento
abandonado	abandonar
abandonados	abandonar

4. Construcción de n-gramas:

En este paso con el texto limpio se dividen las metas en sectores y se construyen n-gramas para n=1,2,3,4. Esto permite la estandarización de las metas por sectores. Además, se generan un archivo de extensión xlsx que contiene las palabras, bigramas, trigramas y cuatrigramas con sus respectivas frecuencias y porcentajes para cada uno de los sectores.

5. **Redacción de metas:** En este es el último paso de acuerdo con los sectores se realiza una revisión y redacción de metas.

5. Resultados

Para cada uno de los sectores se generan los archivos de Excel correspondientes, por ejemplo, para el sector agropecuario se tienen las siguientes palabras, bigramas, trigramas y cuatrigramas.

Palabra	Frecuencia	Porcentaje	Porcentaje acumulado
agropecuario	2633	3.356278	3.356278
productor	2030	2.587635	5.943913
productivo	1919	2.446144	8.390057
rural	1520	1.93754	10.327597
asistencia	1223	1.558955	11.886552
técnico	1193	1.520714	13.407266
producción	1045	1.332059	14.739325
desarrollo	863	1.100064	15.839389
sector	739	0.942001	16.78139
agrícola	682	0.869344	17.650734

Tabla 1. Palabras más frecuentes sector agropecuario

Bigrama	Frecuencia	Porcentaje	Porcentaje acumulado
asistencia técnico	919	1.188091	1.188091
seguridad alimentario	457	0.590813	1.778904
productor agropecuario	367	0.474461	2.253365

sector agropecuario	282	0.364572	2.617937
desarrollo rural	267	0.34518	2.963117
cadena productivo	235	0.30381	3.266927
distrito regar	215	0.277954	3.544881
asociación productor	205	0.265026	3.809907
producción agropecuario	200	0.258562	4.068469
mejoramiento genético	195	0.252098	4.320567

Tabla 2. Bigramas más frecuentes sector agropecuario

Trigrama	Frecuencia	Porcentaje	Porcentaje acumulado
asistencia técnico agropecuario	147	0.192779	0.192779
asistencia técnico integral	113	0.148191	0.34097
consejo desarrollo rural	111	0.145568	0.486538
asistencia técnico productor	99	0.129831	0.616369
asistencia técnico directo	93	0.121962	0.738331
comercialización producto agropecuario	83	0.108848	0.847179
brindar asistencia técnico	77	0.10098	0.948159
beneficiar asistencia técnico	66	0.086554	1.034713
técnico directo rural	62	0.081308	1.116021
servicio asistencia técnico	59	0.077374	1.193395

Tabla 3. Trigramas más frecuentes sector agropecuario

Cuatrigrama	Frecuencia	Porcentaje	Porcentaje acumulado
asistencia técnico directo rural	61	0.081165	0.081165
asociatividad desarrollo empresarial rural	40	0.053223	0.134388
beneficiar asistencia técnico integral	39	0.051892	0.18628
productor rural beneficiar asistencia	38	0.050561	0.236841
rural beneficiar asistencia técnico	37	0.049231	0.286072
través canal financiero formal	35	0.04657	0.332642
productividad reducir costo producción	33	0.043909	0.376551
aumentar productividad reducir costo	31	0.041248	0.417799
asistencia técnico integral pequeño	30	0.039917	0.457716
reducir costo producción agropecuario	30	0.039917	0.497633

Tabla 4. Cuatrigramas más frecuentes sector agropecuario

Se crea un proyecto en R con el nombre de "Terridata_Metas". En la carpeta "Outputs/N-gramas/Graficos" se encuentran los gráficos de barras elaborados para cada uno de los n-gramas y sectores. En la carpeta "Scripts" se encuentra el código utilizado para cada uno de los pasos descritos en la metodología debidamente comentado. Los archivos de Excel completos se encuentran en la carpeta "Outputs/N-gramas".

6. Referencias

Departamento Nacional de Planeación. (2018). *Portal Territorial*. Obtenido de Sistema de Información para la Evaluación de la Eficacia:
<https://portalterritorial.dnp.gov.co/pdt/#/consulta-publica>

Oficina del alto comisionado para la paz. (2018). *PLanes de desarrollo territorial*. Obtenido de Alto Comisionado para la Paz:
<http://www.altocomisionadoparalapaz.gov.co/Documents/informes-especiales/plan-desarrollo-territorial/index.html>