

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**El futuro
es de todos**

DNP
Departamento
Nacional de Planeación



ANÁLISIS DE TERMINOS Y FRASES RELACIONADOS CON TEMAS DE GÉNERO EN LOS PLANES DE DESARROLLO TERRITORIAL 2020 – 2023

Entidad

Departamento Nacional de Planeación

- Dirección de Desarrollo Digital - Unidad de Científicos de Datos
- Dirección de Desarrollo Social – Subdirección de Género

Sector

Planeación

Lenguaje

Python

Fuente de datos

Documentos PDT de los departamentos
Documentos PDT de los municipios
Base códigos DIVIPOLA DANE
Frasas de género identificadas por la Subdirección de Género.

Contenido

1.	Presentación	2
2.	Objetivos del proyecto	2
2.1.	General.....	2
2.2.	Específicos	3
3.	Metodología	3
3.1.	Lectura de los documentos PDT	3
3.2.	Limpieza de texto.....	4
3.3.	Grupos de búsqueda de frases y términos sobre los PDT.....	4
3.4.	Estrategia de búsqueda de frases y términos sobre los PDT	6
3.5.	Indicador para medir la alineación de los PDT con temas de género.....	8
4.	Resultados	9
5.	Conclusiones y recomendaciones	12
6.	Socialización	13
	ANEXOS	14
	Anexo 1 – Expresiones regulares utilizadas en la estrategia de búsqueda Grupo 1	14
	Anexo 2 – Expresiones regulares utilizadas en la estrategia de búsqueda Grupo 2	14
	Anexo 3 – Expresiones regulares utilizadas en la estrategia de búsqueda Grupo 3	20
	Anexo 4 – Municipios por fuera del análisis	21



1. Presentación

Colombia ha venido avanzando en reducir las brechas que existen entre hombres y mujeres en diferentes ámbitos como el del mercado laboral, la participación, el acceso a activos y servicios, entre muchos otros. Sin embargo, aún existen enormes retos para alcanzar la igualdad de género y garantizar los derechos de todas las mujeres. Para lograr este objetivo de igualdad, es fundamental el compromiso de todo el Estado en su conjunto, y de manera esencial la participación de los entes territoriales para generar las condiciones de desarrollo y bienestar de las mujeres en sus propios contextos.

Los Planes de Desarrollo Territorial (PDT) constituyen el instrumento de política más importante para orientar las acciones y apuestas de los entes territoriales en un período de gobierno. El contenido de estos documentos se convierte en la carta de navegación, y por ello resulta fundamental incluir en este instrumento, lineamientos específicos que permitan adelantar las decisiones y acciones necesarias para reducir la brecha entre hombres y mujeres. Por lo anterior, es vital realizar análisis completos de todo lo que se consigna en los PDT con el objetivo de identificar cuáles son las estrategias y falencias territoriales para la formulación de políticas públicas relacionadas en temas de enfoque de género.

Uno de los grandes obstáculos para realizar estos análisis, es la cantidad de información textual presente en estos documentos, pues en Colombia existen 32 departamentos, 1.102 municipios, 18 áreas no municipalizadas y 1 isla, para un total de 1.1153 documentos. Por ende, leer toda esta información de forma manual es una tarea tediosa y que implica un gasto alto en recurso humano y tiempo. Dicho lo anterior y en aras de facilitar esta tarea, la Unidad de Científicos de Datos desarrollo una metodología basada en técnicas de análisis y minería de texto para buscar frases y términos relacionados con e fin de dar un indicador para medir u objetivar que tan alineados están los PDTs departamentales y municipales con los temas de equidad de género.

Colombia has been going forward to progress in reducing the gaps that exist between men and women regarding different areas such as the labor market, involvement spaces, access to assets and services, among many others. However, there are still enormous challenges to achieve gender equality and guarantee the rights of all women. To achieve this, the commitment of the entire State is fundamental, essentially the participation of territorial entities to generate the conditions for the development and well-being of women in their contexts.

The Territorial Development Plans (PDT) constitute the most significant policymaking instrument to guide the actions and challenges of the territorial entities in a period of government. The content of these documents becomes the navigation chart; for this reason, it is essential to include specific guidelines that make it possible to advance the decisions and actions necessary to reduce the gap between men and women. Therefore, it is vital to carry out a complete analysis of everything that is consigned in the PDTs to identify which are the territorial strategies and shortcomings for the public policymaking related to gender issues.

One of the biggest obstacles to carrying out these analyzes is the amount of textual information present in these documents, since, Colombia has 32 departments, 1,102 municipalities, 18 non-municipalized areas, and an island, for a total of 1,1153 documents. Therefore, reading all this information manually is a tedious task that involves a high expenditure on human resources and time. That said, and to facilitate this task, the Unidad de Científicos de Datos developed a methodology based on analysis and text mining techniques to search for phrases and related terms to give an indicator to measure or objectify how aligned are the departmental and municipal PDTs with gender equity issues.

2. Objetivos del proyecto

2.1. General

Identificar frases y términos relacionados con temas de género dentro de los PDT (departamentales y municipales) 2020-2023, para cuantificar que tan alineados están con relación a estos temas mediante el uso de técnicas de minería y análisis de texto.



2.2. Específicos

1. Desarrollar una metodología de análisis y minería de texto que permita identificar frases y términos de género dentro de los PDT.
2. Proponer un indicador de medición de los PDT basados en los términos y frases de género que se encuentren en el documento, de forma que se puedan comparar entre sí
3. Consolidar el análisis para cada uno de los PDT departamentales y municipales con el indicador propuesto.

3. Metodología

3.1. Lectura de los documentos PDT

Los documentos PDT 2020-2023 utilizados como insumo de este proyecto son presentados en formato PDF. Debido a ello, la primera etapa de la metodología es la lectura de cada uno de estos archivos. Esta lectura presenta desafíos especialmente para los archivos que fueron escaneados en baja calidad y subidos como un documento PDF, por ende, en esta etapa de lectura hay dos pasos:

- El primer paso es leer el documento como un archivo de alta calidad, es decir, el computador es capaz de extraer toda la información de texto de forma confiable y con algunos errores mínimos de tipografía y guardarlo como texto plano para posteriormente realizar análisis sobre este. En la Figura 1 se muestra un ejemplo de un fragmento de texto extraído sobre el PDT del municipio de Chinchiná – Caldas, donde el resultado es igual que el texto original.

Figura 1: Texto extraído de documento PDT del municipio de Chinchiná - Caldas

Texto en el documento PDT

Las metas y propuestas contenidas en este Plan de Desarrollo permitirán conducir al municipio hacia el logro de mejores condiciones de vida para sus habitantes; será la carta de navegación para los próximos 4 años de vida institucional y comunitaria.

Texto leído por el computador

Las metas y propuestas contenidas en este Plan de Desarrollo permitirán conducir al municipio hacia el logro de mejores condiciones de vida para sus habitantes; será la carta de navegación para los próximos 4 años de vida institucional y comunitaria.

Fuente: elaboración propia

- El segundo paso se aplica cuando el documento no se puede leer directamente debido a que el texto es un escaneo del documento original, por lo que se debe extraer el texto de las imágenes. Cada uno de estos pasa por una etapa de OCR (*Optical Character Recognition*) para extraer el texto del documento escaneado. En esta etapa, el texto extraído puede contener errores asociados a la calidad del documento y donde el computador no puede reconocer con exactitud el texto desde la imagen. En algunos casos, la calidad del documento es tan baja que el texto que se extrae no es confiable, esto se ilustra en la Figura 2.

Figura 2: Resultado del OCR de fragmento de texto del PDT del municipio de Carmen de Bolívar – Bolívar: El documento es de muy mala calidad.

Texto en el documento PDT

En las materias relacionadas con sus atribuciones, también tienen iniciativa el Personero Municipal y las juntas administradoras locales con jurisdicción en corregimientos o comunas del territorio del municipio.

Texto leído por el computador

.El...byd...bam...mndonmpfayu...hh...
mu...oonmywn...uoo...nh.....doh...yupnntmun
mvmw...aramwd.....yhm"vón.]

Fuente: elaboración propia



3.2. Limpieza de texto

Después de obtener la representación de cada PDT como texto plano, estos pasan por una etapa de limpieza de texto con el fin de remover palabras y caracteres que no son relevantes y que pueden entorpecer el análisis. Por ejemplo, para el computador la palabra “Hola” es diferente de “hola” o de “HOLA”, tener el texto en minúsculas evita este tipo de confusiones, o por ejemplo la palabra “mujer” es diferente de “mujer,” por tener una coma (,) al final de la palabra. Este proceso de limpieza de texto se realiza en tres pasos:

1. Se convierte todo el texto a minúsculas
2. Se remueven signos de puntuación, tales como comas, puntos, punto y coma; y se remueven espacios múltiples.
3. Por último, se quitan nombres de hombres y mujeres, nombres de departamentos y municipios y números presentes del documento

En la Figura 3 se muestra un ejemplo del resultado de la etapa de limpieza de texto sobre un fragmento extraído del PDT de Medellín – Antioquia.

Figura 3: Ejemplo de la etapa de limpieza de texto

Texto original

El Plan de Desarrollo de acuerdo con la Ley 152 de 1994 el componente estratégico se desagrega de la siguiente manera: además de una introducción, la propuesta estratégica, el diagnóstico y el marco programático.

Texto después de la etapa de limpieza

el plan de desarrollo de acuerdo con la ley de el componente estratégico se desagrega de la siguiente manera además de introducción la propuesta estratégica el diagnóstico y el marco programático

Fuente: elaboración propia

3.3. Grupos de búsqueda de frases y términos sobre los PDT

Con el objetivo de medir que tan alineados están los PDT con temas de género, se propuso una estrategia de búsqueda basada en 3 grupos de análisis, propuestos por la subdirección de género. En el primer grupo se buscan 14 conceptos generales de género, consignados en la Tabla 1, con el fin de tener una visión a grandes rasgos de las estrategias de cada ente territorial con respecto a los temas de género.

Tabla 1: Conceptos generales sobre género

Conceptos generales			
1	Acciones afirmativas hacia las mujeres	8	Enfoque de género
2	Brechas de género	9	Equidad de género para las mujeres
3	Brechas entre mujeres y hombres	10	Mujeres en el desarrollo
4	Derechos de las mujeres	11	Oportunidades para las mujeres
5	Desigualdad de género	12	Perspectiva de género



6	Discriminación contra las mujeres	13	Situación de desventaja de las mujeres
7	Empoderamiento de las mujeres	14	Situación política, social, económica o cultural de las mujeres

Fuente: elaboración propia

En el segundo grupo, se buscaron 207 frases de género agrupadas en 7 categorías y 27 subcategorías (ver Tabla 2). Para cada una de estas subcategorías se tiene en cuenta las posibles combinaciones con un conjunto de términos de enfoque de género habilitadores, es decir, para considerar que una frase asociada a una subcategoría está presente en el texto, es obligatorio que en el mismo fragmento esté presente por lo menos uno de estos términos habilitadores. El conjunto de términos habilitadores se presenta en la Tabla 3.

Tabla 2: Ejes categóricos de búsqueda de temas de género.

Categoría	Subcategoría
Autonomía económica y acceso a activos	Acceso a activos rurales
	Contratación y compra pública
	Economía del cuidado
	Empleo
	Empoderamiento
	Emprendimiento y fortalecimiento empresarial
	Formación para el trabajo
	Inclusión Financiera
	Indicadores de diagnóstico y/o seguimiento
	Medio ambiente
	Vivienda
Vulnerabilidad	
Construcción de paz	Construcción de paz
Educación y acceso a TICS	Áreas de desempeño
	Educación básica, media y superior
	TICS
Institucionalidad y transformación cultural	Fortalecimiento institucional
	Planeación y política
	Transformación cultural
Participación en espacios de toma de decisión	Organizaciones
	Participación en espacios de decisión
	Participación política
Salud y derechos sexuales y reproductivos	derechos sexuales y reproductivos
	Salud
Vida libre de violencias	Acciones de prevención, atención y acceso a justicia
	otras prácticas relacionadas con la violencia
	Violencias

Fuente: elaboración propia



Tabla 3: Términos de enfoque habilitadores

Términos de enfoque		
Mujer	Mujeres	Femenino
Femenina	Género	Feminización

Fuente: elaboración propia

Por último, en el grupo 3, se buscan términos transversales – intersectoriales de temas de género reportados en la Tabla 4. Para este grupo también es necesario que vayan acompañados en el mismo fragmento de texto por un término habilitador de la Tabla 3.

Tabla 4: Términos transversales – intersectoriales

Transversales – intersectoriales		
Rural	Joven	Adulta
Niña	Adolescente	LBTI
Lesbiana	Trans	Indígena
Negra	Afrodescendiente	Raizal
Palenquera	NARP	Room
Étnico	Víctima	Discapacidad

Fuente: elaboración propia

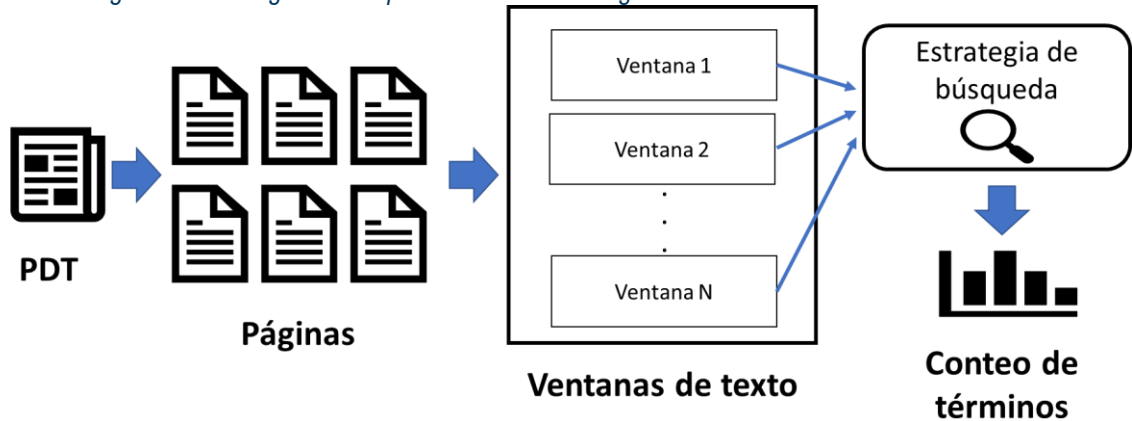
Dada la condición de los términos habilitadores, si en un fragmento de texto aparece la palabra “rural”, para ser considerado un término relevante para el análisis debe ir acompañado de alguno de los términos de la Tabla 3. Es decir, si aparece la palabra “mujer”, es probable que en ese fragmento se mencione “mujer rural”.

3.4. Estrategia de búsqueda de frases y términos sobre los PDT

En la Figura 4, se muestra el esquema general de la búsqueda de término relacionados con género en documentos PDT. Este esquema se aplica a los tres grupos de búsqueda mencionados en la Sección 3.3. Cada documento PDT después de la etapa de limpieza de texto, pasa por los siguientes pasos:

1. El documento es dividido en páginas para su análisis de términos.
2. Cada página del PDT a su vez es dividido en fragmentos de texto más pequeños, denominados ventanas de análisis, es decir, pueden existir varias ventanas por página. Cada ventana de análisis está compuesta por 200 palabras (tamaño promedio de un párrafo), debido a que es difícil identificar correctamente párrafos en el texto. En algunos casos puede contener menos palabras debido al número de palabras presentes por página, es decir, una página con 453 palabras tendrá entonces 2 ventanas de 200 palabras y una de 53 palabras.
3. Cada ventana entra a la estrategia de búsqueda de términos relevantes y al final del proceso se consolida el conteo de número de apariciones de los términos en las ventanas. La estrategia de búsqueda se explica a continuación:

Figura 4: Estrategia de búsqueda de términos de género en documentos PDT



Fuente: elaboración propia

Debido a que una frase de género puede aparecer expresada de muchas maneras, la estrategia de búsqueda utiliza una metodología basada en expresiones regulares¹ para cada frase de los grupos de términos de interés. Por ejemplo, la frase “brechas entre hombres y mujeres” puede aparecer como “brechas entre mujeres y hombres”; “brecha entre hombre y mujer” o “brecha de género entre hombre y mujer”, entre muchas otras combinaciones. Tratar de armar todas las combinaciones posibles sobre el cómo aparece la frase, puede dar como resultado cientos de frases. Por tal motivo, las expresiones regulares dan mecanismos para hacer esa búsqueda más lapsa pudiendo identificar múltiples formas de la frase. Para dar solución al problema anterior se utiliza la expresión regular “brechas*|mujer(?:es)*|hombres*”. En la Figura 5 se ilustra su funcionamiento sobre un texto de prueba; la expresión “brechas*” es capaz de encontrar las palabras brecha o brechas; la expresión “mujer(?:es)*” es capaz de encontrar las palabras mujer y mujeres; por último, la expresión *hombres** encuentra las palabras hombre u hombres. La raya vertical o pleca (|) actúa como una condición lógica “o” y permite que se busquen las tres expresiones al mismo tiempo.

Figura 5: Ejemplo del funcionamiento de las expresiones regulares

Expresión regular : **brechas*|mujer(?:es)*|hombres***

las brechas que existen entre hombres y mujeres es igual a la brecha entre hombre y mujer o brecha entre mujer y hombre

Texto de prueba

Brechas entre hombres y mujeres: 3 ocurrencias

Fuente: elaboración propia

Después de identificar todas las palabras que cumplen el patrón de la expresión regular, se hace un conteo por cada condición separada por la pleca (|). En el ejemplo de la Figura 5, hay 3 condiciones y cada una tiene 3 ocurrencias.

¹ Las expresiones regulares son patrones que se utilizan para hacer coincidir combinaciones de caracteres en cadenas de texto.



Para encontrar el número de apariciones de la frase “brechas entre hombres y mujeres”, se halla el mínimo de ocurrencias entre todas las condiciones, en este caso 3, por lo que la estrategia de búsqueda da como resultado 3 para esa ventana de análisis.

Cada una de las frases y términos de género de los 3 grupos de análisis tienen asociada una expresión regular para la estrategia de búsqueda (ver Anexo 1, Anexo 2, Anexo 3). Este proceso se hace para cada término y para cada ventana dentro del PDT. Al final se consolidan los conteos de los términos para todo el documento. Se aclara, que para los grupos de búsqueda de ejes categóricos (grupo 2, ver Tabla 2) y términos transversales – intersectoriales (grupo 3, ver Tabla 4), a la expresión regular de búsqueda se le agrega las condiciones de los temas de enfoque (ver Tabla 3). Esto hace que la condición de búsqueda sea válida, si y solo si, está presente en la ventana al menos uno de los términos habilitadores.

3.5. Indicador para medir la alineación de los PDT con temas de género

Después de consolidar los conteos por cada frase o término asociadas a los grupos de búsqueda, se tiene un total de 59 dimensiones (14 generales, 27 ejes categóricos y 18 transversales). A partir de estas 59 dimensiones, se proponen dos puntajes para la comparación de los PDT.

- **Promedio de términos**

El promedio de términos es una sumatoria sobre el número de ocurrencias de las 59 dimensiones en un PDT, normalizado por el número de páginas (con información, se ignoran páginas con menos de 5 palabras) que contenga el documento PDT. Este indicador es la forma natural de medir la aparición de términos de género en el documento. Matemáticamente, el indicador I_i sobre el i -ésimo documento PDT se define como:

$$I_i = \frac{\sum_v \sum_{j=1}^{59} d_v(j)}{P_i}$$

Donde el subíndice v representa la ventana de análisis sobre el texto, $d(j)$ representa el número de ocurrencias de la dimensión j en la ventana v y P_i es el número de páginas con información del i -ésimo PDT.

- **Promedio ponderado de términos**

Esta medida se plantea con el objetivo de conceder más importancia a las ventanas que contienen más de una dimensión durante la búsqueda, bajo el supuesto que los documentos PDT que tengan capítulos dedicados a temas de género obtengan una mayor puntuación en comparación a los que mencionan de forma aislada dichos temas. Matemáticamente, el indicador promedio ponderado IP_i sobre el i -ésimo documento PDT se define como:

$$IP = \frac{\sum_v \sum_{j=1}^{59} D_v d_v(j)}{P_i}$$

Donde el subíndice v representa la ventana de análisis sobre el texto, $d(j)$ representa el número de ocurrencias de la dimensión j en la ventana v , D_v representa el número de dimensiones diferentes que aparecen sobre la ventana v y P_i es el número de páginas con información del i -ésimo PDT.

Para dar un entendimiento más profundo de la diferencia entre estos dos indicadores se da el siguiente ejemplo:

Suponga que se tiene un documento PDT con 10 páginas, en el cual se analizan 3 ventanas, como se consigna en la Tabla 5



Tabla 5: Resultados para ejemplificar los indicadores

páginas en el documento: 10		
Ventanas	Dimensiones encontradas	Conteo
1	Brecha de género	1
	Desigualdad de género	2
	Empoderamiento de las mujeres	1
2	Enfoque de género	1
3	Brecha de género	3
	Enfoque de género	1

Fuente: elaboración propia

El indicador de promedio de términos hace la sumatoria sobre todas las ventanas y dimensiones (9 ocurrencias) y lo divide entre el total de páginas (10 páginas) para un puntaje de 0,9.

El indicador ponderado, primero hace la suma de ocurrencias sobre cada una de las ventanas (4 en la primera, 1 en la segunda y 4 en la tercera), después se fija cuantas dimensiones aparecen en cada ventana (3 dimensiones en la primera, 1 dimensión en la segunda, 2 dimensiones en la tercera). Por último, el número de dimensiones de cada ventana por el número de ocurrencias de términos, esto es, $4 * 3$ para la primera ventana, $1*1$ para la segunda y $4*2$ para la tercera. Se hace la suma de estos valores y se divide sobre el número de páginas. El resultado de este indicador será $21/10 = 2,1$.

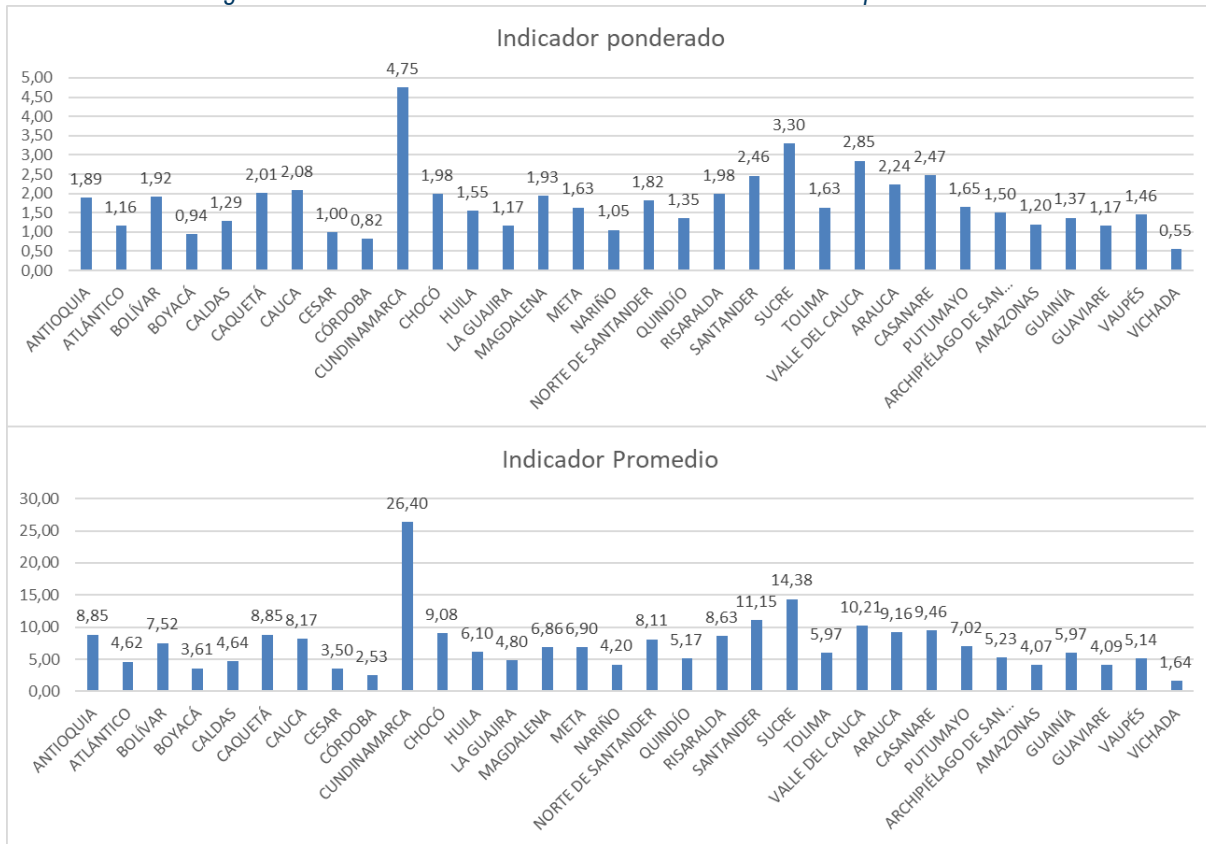
La interpretación de los indicadores propuestos se debe realizar en un contexto global, pues el valor por si solo no provee información relevante. Visto en conjunto, expresan de manera general cuantas frases o términos de temas de género fueron encontradas en los documentos analizados.

4. Resultados

La metodología propuesta para búsqueda de temas de género e índice de medición se aplicó de forma automática a los 32 documentos PDT 2020 -2023 de los entes departamentales. La Figura 6 muestra los resultados obtenidos después de la estrategia de búsqueda sobre los documentos y el comportamiento de los dos indicadores propuestos. Aunque el indicador ponderado acentúa las diferencias entre los documentos, la tendencia de los dos indicadores es la misma.



Figura 6: Resultados de los dos indicadores sobre los PDT departamentales



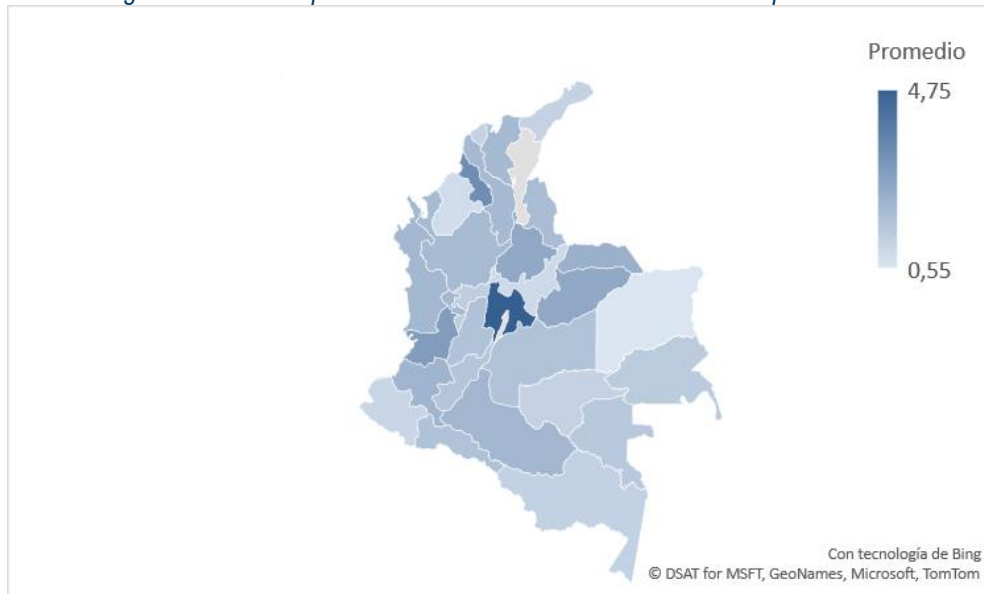
Fuente: elaboración propia

En la Figura 7, se muestra el indicador promedio calculado a partir del análisis de los documentos PDT a nivel departamental. No es necesario graficar el indicador ponderado, debido que sigue la misma tendencia como se mencionó anteriormente.

A partir de los resultados obtenidos, se visualiza que los departamentos que mayor relevancia le dieron a los temas de género fueron Cundinamarca, seguido de Sucre y Valle del Cauca. Los PDT en los que menos términos relacionados con temas de género fueron Boyacá, Córdoba y Vichada.



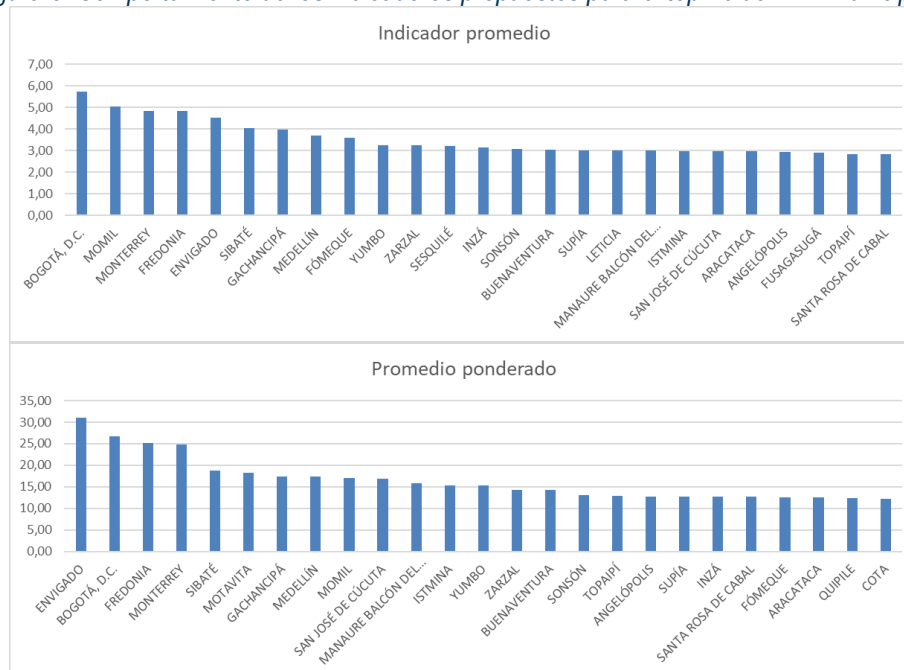
Figura 7: Indicador promedio del análisis de los PDT a nivel departamental



Fuente: elaboración propia

A diferencia de los PDT departamentales, el indicador promedio y el indicador ponderado dan puntuaciones que difieren en que tan alineado están estos documentos con los temas de género. En la Figura 8, se muestra el comportamiento de los dos indicadores propuestos sobre los 26 municipios con puntaje más alto. Se destaca el municipio de Fômeque – Cundinamarca (este documento presentó muchas fallas en la etapa de OCR) que cuenta con 10 páginas, el indicador de promedio se ve sesgado por esa cantidad pequeña, mientras que el indicador ponderado penaliza de mejor manera a este municipio.

Figura 8: Comportamiento de los indicadores propuestos para el top 26 de PDT municipales

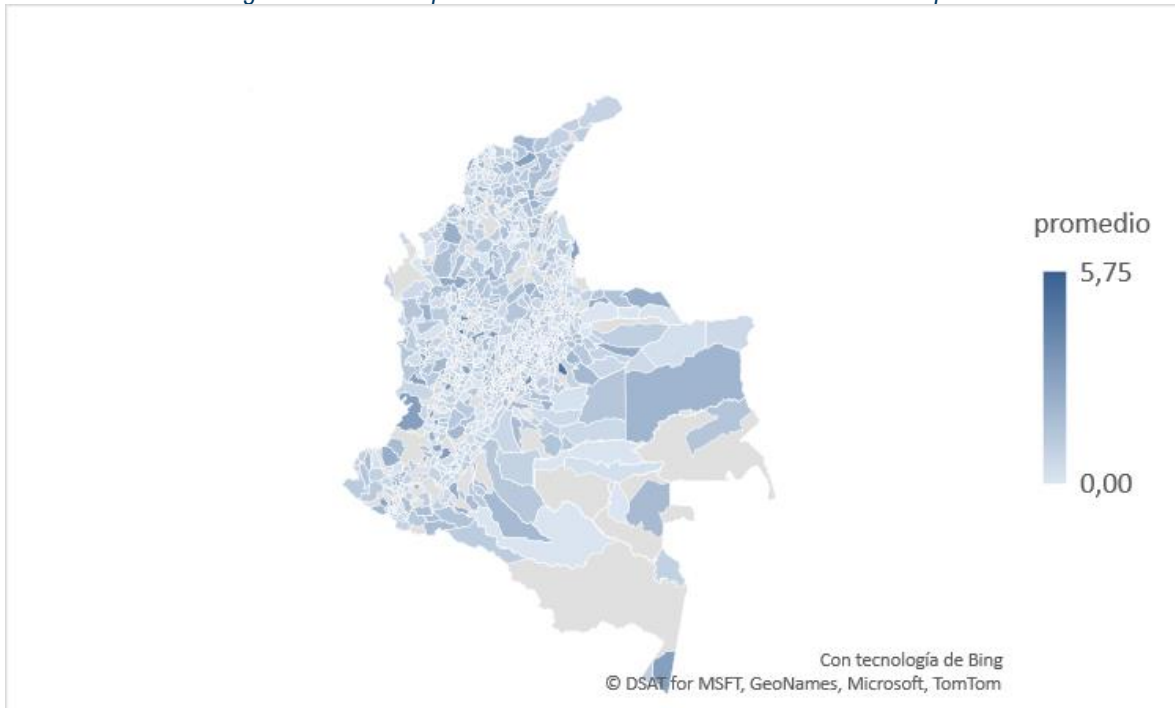


Fuente: elaboración propia



También se presenta la visualización del indicador promedio para todos los municipios de los que fue posible obtener el PDT (ver Figura 9, en total 1075 municipios. Los municipios que se quedaron fuera de este análisis se reportan en la sección de ANEXOS, Anexo 4.

Figura 9: Indicador promedio del análisis de los PDT a nivel municipal



Fuente: elaboración propia

5. Conclusiones y recomendaciones

A partir de la metodología implementada y de los resultados obtenidos para cumplir los objetivos de este proyecto, planteados en el plan de trabajo, se presentan a continuación las principales conclusiones obtenidas por el equipo de la UCD y las principales recomendaciones para un mejor uso y aprovechamiento del proyecto.

1. Se desarrolló una metodología de búsqueda de términos relacionados con temas de género basado en expresiones regulares, lo que permite encontrar múltiples formas de una frase en los textos.
2. Se entregaron dos hojas de cálculo con los resultados consolidados para departamentos y municipios, donde se puede consultar los indicadores promedio y ponderado, y el conteo de frases encontradas en cada documento PDT.
3. Debido a la cantidad de documentos a nivel municipal que se analizaron, no fue posible identificar cuáles de ellos no presentan una calidad aceptable de la extracción de texto a partir de OCR. Por lo que se recomienda excluir de los análisis aquellos que presenten puntajes de cero o muy cercanos a cero en el indicador de promedio.
4. Debido a que este fue un análisis automatizado sobre los textos de los PDT, se aconseja que los resultados se analicen desde la visión de un experto temático.
5. Por ahora los dos indicadores propuestos son susceptibles a documentos que presentan pocas páginas con información relevante (puntajes muy altos) como el caso de Fómeque o a documentos donde la etapa de OCR falló, pues presentan muchas páginas, pero pocos términos relacionados (puntajes muy bajos). Para ello se recomienda como trabajo futuro, identificar los PDT en los que se presenta este problema.



6. Como trabajo futuro de aconseja trabajar en conjunto con los expertos temáticos para proponer un indicador más robusto y que brinde mayor información, debido a que los actuales, simplemente cuantifican la cantidad de frases presentes en los documentos PDT.

6. Socialización

Este proyecto se socializó con la Subdirección de Género y la Unidad de Científicos de Datos.



ANEXOS

Anexo 1 – Expresiones regulares utilizadas en la estrategia de búsqueda Grupo 1

Tabla 6: Expresiones regulares para grupo 1 – temas generales de género

Grupo 1: términos generales	
Frase de género	expresión regular
Acciones afirmativas hacia las mujeres	acci[ó]n(?:es)* afirmativas* mujer
Brechas de género	brecha género
Brechas entre mujeres y hombres	brecha mujer hombre
Derechos de las mujeres	derecho mujer
Desigualdad de género	igualda género
Discriminación contra las mujeres	discrimina mujer
Empoderamiento de las mujeres	empodera mujer
Enfoque de género	enfoque género
Equidad de género para las mujeres	equidad género mujer
Mujeres en el desarrollo	desarrollo mujer
Oportunidades para las mujeres	oportunidad mujer
Perspectiva de género	perspectiva género
Situación de desventaja de las mujeres	desventaja mujer

Fuente: elaboración propia

Anexo 2 – Expresiones regulares utilizadas en la estrategia de búsqueda Grupo 2

Tabla 7: Expresiones regulares para grupo 2 – temas categóricos de género

concepto	Frase de género	Expresión regular
Acceso a activos rurales	insumos agrícolas	insumo agr[ri]cola
Acceso a activos rurales	maquinaria	maquinaria
Acceso a activos rurales	propiedad tierra	propiedad tierra
Contratación y compra pública	compra pública	compra p[ú]blica
Contratación y compra pública	contrato público	contrato p[ú]blico
Economía del cuidado	cuidadoras	cuidadora
Economía del cuidado	labores de cuidado	labor cuidado
Economía del cuidado	labores domésticas	labor dom[é]stic[ao]s
Economía del cuidado	oficios del hogar	oficio hogar
Economía del cuidado	trabajo de cuidado no remunerado	trabaj[ao] cuidado [sn][io]n* remunera
Empleo	empleabilidad	empleabilidad
Empleo	empleada	empleada
Empleo	empleo	empleo
Empleo	igualdad salarial	igualdad salari[ao]*



Empleo	incursión laboral	incers[i]o[n] laboral
Empleo	inclusión laboral	inclu[is][ri](?:[oó]n)* laboral
Empleo	inclusión productiva	inclu[is][ri](?:[oó]n)* productiva
Empleo	oportunidad laboral	oportunidad laboral
Empleo	ruta de empleo	ruta empleo
Empleo	trabajo decente	trabajo decente
Empoderamiento	autonomía económica	autonom[ii]a econ[oó]mica
Empoderamiento	empoderamiento económico	empodera econ[oó]mic[ao]
Empoderamiento	impulso económico	impuls[ao]r* econ[oó]mic[ao](?:.mente)*
Empoderamiento	seguridad económica	seguridad econ[oó]mica
Emprendimiento y fortalecimiento empresarial	asociatividad	asociatividad
Emprendimiento y fortalecimiento empresarial	competitividad	competitiv[ai](?:.dad)*
Emprendimiento y fortalecimiento empresarial	emprededora	emprededora
Emprendimiento y fortalecimiento empresarial	emprendimiento	emprendimiento
Emprendimiento y fortalecimiento empresarial	empresa	empresa
Emprendimiento y fortalecimiento empresarial	encadenamiento productivo	encadenamiento productivo
Emprendimiento y fortalecimiento empresarial	iniciativa económica	iniciativa econ[oó]mica
Emprendimiento y fortalecimiento empresarial	iniciativa laboral	iniciativa laboral
Emprendimiento y fortalecimiento empresarial	iniciativa productiva	iniciativa productiva
Emprendimiento y fortalecimiento empresarial	negocio	negocio
Emprendimiento y fortalecimiento empresarial	productividad	productividad
Emprendimiento y fortalecimiento empresarial	proyecto productivo	proyecto productivo
Emprendimiento y fortalecimiento empresarial	red de comercialización	red comercial
Formación para el trabajo	asistencia técnica económica	asist[eij][nr] técnic[ao] econ[oó]mic[ao]
Formación para el trabajo	capacitación	capacitaci[oó]n
Formación para el trabajo	educación para trabajo	educaci[oó]n trabajo
Formación para el trabajo	extensión	extensi[oó]n
Formación para el trabajo	formación para trabajo	forma[rc](?:.i[oó]n)* trabaj[ao]r*
Formación para el trabajo	habilidades	habilidad
Formación para el trabajo	proceso formativo	proceso formativo
Inclusión Financiera	bancarización	bancarizaci[oó]n
Inclusión Financiera	crédito	cr[eé]dito



Inclusión Financiera	educación financiera	educaci[oo]n financiera
Inclusión Financiera	financiación	financia[rc](?:i[oo]n)*
Inclusión Financiera	inclusión financiera	inclu[is][ri](?:[oo]n)*financiera
Inclusión Financiera	servicios financieros	servicio financiero
Indicadores de diagnóstico y/o seguimiento	brecha salarial	brecha salar[oa]
Indicadores de diagnóstico y/o seguimiento	cabeza de familia	cabeza familia
Indicadores de diagnóstico y/o seguimiento	cuidado no remunerado	cuidado no remunerad[ao]
Indicadores de diagnóstico y/o seguimiento	economía del cuidado	econom[í]a del cuidado
Indicadores de diagnóstico y/o seguimiento	informalidad	informal
Indicadores de diagnóstico y/o seguimiento	ingresos propios	ingresos* propios*
Indicadores de diagnóstico y/o seguimiento	jefe de hogar	jefe hogar
Indicadores de diagnóstico y/o seguimiento	ocupación	ocupaci[oo]n
Indicadores de diagnóstico y/o seguimiento	pobreza	pobreza
Indicadores de diagnóstico y/o seguimiento	pobreza en el tiempo	pobreza tiempo
Indicadores de diagnóstico y/o seguimiento	tasa de desempleo	tasa empleo
Indicadores de diagnóstico y/o seguimiento	tasa de participación	tasa participa[cr](?:i[oo]n)*
Medio ambiente	biodiversidad	biodiversidad
Medio ambiente	cambio climático	cambio clim[aa]tico
Medio ambiente	impacto ambiental	impact[ao]r* ambient[ae]]*
Vivienda	compra vivienda	compra vivienda
Vivienda	mejoramiento vivienda	mejora vivienda
Vivienda	subsidio vivienda	subsidi[ao]r* vivienda
vulnerabilidad	apoyo económico	apoy[ao] económic[ao]
vulnerabilidad	subsidio	subsidi[ao]r*
vulnerabilidad	transferencia monetaria	transferencia monetaria
Construcción de paz	desplazamiento	desplazamiento
Construcción de paz	dignificación	dignificaci[oo]n
Construcción de paz	memoria histórica	memoria hist[oo]rica
Construcción de paz	reparación integral	repara[rc]* integral
Construcción de paz	reparación simbólica	repara[rc]* simb[oo]lica
Construcción de paz	restitución de tierras	restitu[ic] tierra
Áreas de desempeño	áreas de desempeño no tradicional	[aa]rea desempeño no tradicional



Áreas de desempeño	ciencia	ciencia
Áreas de desempeño	diversificación educativa	diversificaci[oó]n educativ[ao]
Áreas de desempeño	diversificación ocupacional	diversificaci[oó]n ocupaci[oó]n
Áreas de desempeño	diversificación profesional	diversificaci[oó]n profesional
Áreas de desempeño	ingeniería	ingenier[ii]a
Áreas de desempeño	matemáticas	matem[aa]ticas
Áreas de desempeño	stem	\bstem\b
Áreas de desempeño	tecnología	tecnolog[ii]a
Educación básica, media y superior	analfabeta	analfabeta
Educación básica, media y superior	aprendizaje igualitario	aprendizaje igualitari[ao]
Educación básica, media y superior	deserción	deserci[oó]n
Educación básica, media y superior	educación	educaci[oó]n
Educación básica, media y superior	educación inclusiva	educaci[oó]n inclusiva
Educación básica, media y superior	educación media	educaci[oó]n media
Educación básica, media y superior	educación superior	educaci[oó]n superior
Educación básica, media y superior	educación técnica y tecnológica	educaci[oó]n t[eé]cn[io]cl[aoó]
Educación básica, media y superior	nivel educativo	nivel educa[ct]i
Educación básica, media y superior	permanencia en el sistema educativo	permane[cn][ce] sistema educa[ct]i[voó]*
Educación básica, media y superior	permanencia escolar	permane[cn][ce] escolar
Educación básica, media y superior	resultados saber	resultados* saber
TICS	acceso a internet	acce[ds][eo] internet
TICS	alfabetización digital	alfabetizaci[oó]n digital
TICS	celular	celular
TICS	computador	computador
TICS	desarrollo digital	desarrollo digital
TICS	emprendimiento digital	emprendimiento digital
TICS	formación digital	forma[rc](?:i[oó]n)* digital
TICS	inteligencia artificial	inteligencia artificial
TICS	plataforma digital	plataforma digital
TICS	puntos digitales	punto digital
TICS	tablet	tablet
TICS	telefonía móvil	telefon[oi]a* m[oó]vil\s
TICS	uso y apropiación	uso apropiaci[oó]n
Fortalecimiento institucional	casa mujer	casa mujer
Fortalecimiento institucional	comisión de equidad	comisi[oó]n equidad
Fortalecimiento institucional	fortalecimiento de enlaces	fortale[rc](?:imiento)* enlace
Fortalecimiento institucional	fortalecimiento de la oficina	fortale[rc](?:imiento)* oficina
Fortalecimiento institucional	fortalecimiento de la secretaría	fortale[rc](?:imiento)* secretaría
Fortalecimiento institucional	mecanismo de adelanto	mecanismo aselanto



Planeación y política	observatorio	observatorio
Planeación y política	plan de transversalización	plan transversalizaci[oo]n
Planeación y política	política pública	pol[i]tica pública
Planeación y política	presupuesto con enfoque de género	presupuesto enfoque género
Planeación y política	trazador presupuestal	trazador presupuestal
Transformación cultural	cambio cultural	cambio cultural
Transformación cultural	estereotipo	estereotipo
Transformación cultural	estigmatización	estigmatiza
Transformación cultural	imaginarios	imaginario
Transformación cultural	lenguaje incluyente	lenguaje incluyente
Transformación cultural	machismo	mach[ii]smo
Transformación cultural	roles	\brol(?:es)*\b
Transformación cultural	sensibilización de funcionarios	sensibiliza[cr] funcionario
Transformación cultural	sesgo	sesgo
Organizaciones	agremiaciones	agremiaci[oo]n
Organizaciones	asociaciones	asociaci[oo]n
Organizaciones	movimiento social	movimiento social
Organizaciones	organización	organizaci[oo]n
Organizaciones	veeduría	veedur[ii]a
Participación en espacios de decisión	junta de acción comunal	junta acci[oo]n comunal
Participación en espacios de decisión	participación comunitaria	participa[cr](?:i[oo]n)* comuni[dt]a[dr](ia)*
Participación en espacios de decisión	participación en empresa	participa[cr](?:i[oo]n)* empresa
Participación en espacios de decisión	participación en junta directiva	participa[cr](?:i[oo]n)* juntas* directiva
Participación en espacios de decisión	participación social	participa[cr](?:i[oo]n)* social
Participación política	candidata	candidata
Participación política	cargo de elección popular	cargo elecci[oo]n popular
Participación política	cuota	cuota
Participación política	escuela de liderazgo político	escuela liderazgo pol[ii]tic[ao]
Participación política	escuela itinerante	escuela itinerante
Participación política	formación política	forma[rc](?:i[oo]n)* pol[ii]tic[ao]
Participación política	gabinete paritario	gabinete paritario
Participación política	lideresa	lideresa
Participación política	nivel decisorio	nivel decisorio
Participación política	participación política	participa[cr](?:i[oo]n)* pol[i]tica
Participación política	protección de lideresas	prote[cg][ce] lideresa
derechos sexuales y reproductivos	derechos sexuales	derecho sexual
derechos sexuales y reproductivos	Educación para la sexualidad	educaci[oo]n sexual
derechos sexuales y reproductivos	embarazo adolescente	embarazo adolescente
derechos sexuales y reproductivos	embarazo de alto riesgo	embarazo alto riesgo



derechos sexuales y reproductivos	fecundidad adolescente	fecundidad adolescente
derechos sexuales y reproductivos	métodos anticonceptivos	m[eé]todo anticoncep
derechos sexuales y reproductivos	salud reproductiva	salud reproductiva
Salud	atención diferencial en salud	atenci[oó]n diferencial salud
Salud	cancer cuello uterino	cancer cuello uterino
Salud	cancer de mama	cancer mama
Salud	depresión	depresi[oó]n
Salud	enfermedad prevalente	enfermedad prevalente
Salud	infraestructura de salud	infraestructura salud
Salud	mortalidad materna	mortalidad materna
Salud	personal de la salud	persona salud
Salud	salud integral	salud integral
Salud	salud mental	salud mental
Salud	servicios de spa	servicio \bspa\b
Salud	suicidio	suicidio
Salud	VIH	\bvih\b
Salud	VPH	\bvph\b
Acciones de prevención, atención y acceso a justicia	abordaje intersectorial de violencia	aborda[jr] intersectorial violencia
Acciones de prevención, atención y acceso a justicia	acceso a justicia	acce[ds][eo] justicia
Acciones de prevención, atención y acceso a justicia	atención de violencia	atenci[oó]n violencia
Acciones de prevención, atención y acceso a justicia	autonomía física	autonom[ii]a ff[ii]sica
Acciones de prevención, atención y acceso a justicia	casa de acogida	casas* de acogida
Acciones de prevención, atención y acceso a justicia	comisaría familia	comisar[ii]a familia
Acciones de prevención, atención y acceso a justicia	masculinidad corresponsable	masculinidad corresponsable
Acciones de prevención, atención y acceso a justicia	masculinidad no violenta	masculinidad [sn][io]n* violen
Acciones de prevención, atención y acceso a justicia	mesa intersectorial de violencia	mesa intersectorial violencia
Acciones de prevención, atención y acceso a justicia	nuevas masculinidades	nueva masculinidad
Acciones de prevención, atención y acceso a justicia	prevención violencia	preven[ic][ir] violencia
Acciones de prevención, atención y acceso a justicia	registro de casos de violencia	registr[ao] caso violencia
Acciones de prevención, atención y acceso a justicia	sensibilización de jueces	sensibiliza[cr] jue[zc](?:es)*
Acciones de prevención, atención y acceso a justicia	sensibilización personal salud	sensibiliza[cr] persona salud



Acciones de prevención, atención y acceso a justicia	consejo de seguridad	consejo seguridad
otras prácticas relacionadas con la violencia	matrimonio infantil	matrimonio infantil
otras prácticas relacionadas con la violencia	mutilación genital	mutilaci[oó]n genital
otras prácticas relacionadas con la violencia	unión temprana	unión temprana
violencias	acoso laboral	acos[ao] laboral
violencias	acoso sexual	acos[ao] sexual
violencias	amenaza	amenaza
violencias	ataque	ataque
violencias	explotación sexual	explotaci[oó]n sexual
violencias	feminicidio	feminicidio
violencias	homicidio	homicidio
violencias	trata de personas	trata de personas
violencias	víctima de violencia	v[íi]ctima violencia
violencias	violencia de género	violencia género
violencias	violencia de pareja	violencia pareja
violencias	violencia económica	violencia econ[oó]mica
violencias	violencia escolar	violencia escolar
violencias	violencia intrafamiliar	violencia intrafamiliar
violencias	violencia política	violencia política
violencias	violencia sexual	violencia sexual

Fuente: elaboración propia

Anexo 3 – Expresiones regulares utilizadas en la estrategia de búsqueda Grupo 3

Tabla 8: Términos transversales – intersectoriales

Frase de género	Expresión Regular
adolescente	Adolescente
adulto	Adulto
afrodescendiente	Afrodescendiente
discapacidad	Discapacidad
étnico	[eé]tnico
indígena	ind[íi]gena
jóven	j[oó]ven
lbt	Lbt
lesbiana	Lesbiana
NARP	\bnarp\b



negra	Negra
niña	\bni[nñ]as*b
palenquera	palenquera
raizal	Raizal
ROOM	\broom\b
rural	rural
trans	\btrans\b
víctima	v[ij]ctima

Fuente: elaboración propia

Anexo 4 – Municipios por fuera del análisis

Tabla 9: Municipios que se quedaron por fuera del análisis debido a que no se encontró PDT

cod_dpto	Departamento	cod_mpio	Municipio
05	ANTIOQUIA	05145	CARAMANTA
05	ANTIOQUIA	05664	SAN PEDRO DE LOS MILAGROS
05	ANTIOQUIA	05679	SANTA BÁRBARA
05	ANTIOQUIA	05045	APARTADÓ
13	BOLÍVAR	13580	REGIDOR
15	BOYACÁ	15104	BOYACÁ
15	BOYACÁ	15232	CHÍQUIZA
15	BOYACÁ	15696	SANTA SOFÍA
15	BOYACÁ	15757	SOCHA
23	CÓRDOBA	23350	LA APARTADA
25	CUNDINAMARCA	25580	PULÍ
25	CUNDINAMARCA	25867	VIANÍ
25	CUNDINAMARCA	25781	SUTATAUSA
25	CUNDINAMARCA	25281	FOSCA
27	CHOCÓ	27800	UNGUÍA
47	MAGDALENA	47703	SAN ZENÓN
50	META	50689	SAN MARTÍN
52	NARIÑO	52240	CHACHAGÜÍ
68	SANTANDER	68575	PUERTO WILCHES
68	SANTANDER	68179	CHIPATÁ
70	SUCRE	70678	SAN BENITO ABAD
73	TOLIMA	73504	ORTEGA
76	VALLE DEL CAUCA	76036	ANDALUCÍA
76	VALLE DEL CAUCA	76126	CALIMA
76	VALLE DEL CAUCA	76041	ANSERMANUEVO



85	CASANARE	85125	HATO COROZAL
88	ARCHIPIÉLAGO DE SAN ANDRÉS, PROVIDENCIA Y SANTA CATALINA	88001	SAN ANDRÉS
91	AMAZONAS	91460	MIRITÍ - PARANÁ
91	AMAZONAS	91405	LA CHORRERA
91	AMAZONAS	91669	PUERTO SANTANDER
91	AMAZONAS	91536	PUERTO ARICA
91	AMAZONAS	91530	PUERTO ALEGRÍA
91	AMAZONAS	91407	LA PEDRERA
91	AMAZONAS	91263	EL ENCANTO
91	AMAZONAS	91430	LA VICTORIA
91	AMAZONAS	91798	TARAPACÁ
94	GUAINÍA	94885	LA GUADALUPE
94	GUAINÍA	94883	SAN FELIPE
94	GUAINÍA	94888	MORICHAL
94	GUAINÍA	94886	CACAHUAL
94	GUAINÍA	94887	PANA PANA
94	GUAINÍA	94884	PUERTO COLOMBIA
94	GUAINÍA	94343	BARRANCOMINAS
97	VAUPÉS	97511	PACOA
97	VAUPÉS	97889	YAVARATÉ
97	VAUPÉS	97777	PAPUNAHUA

Fuente: elaboración propia