

Dirección de Desarrollo Digital

Unidad de Científicos
de Datos



**DEPARTAMENTO
NACIONAL DE PLANEACIÓN**



MEDICIÓN DEL USO DE EVALUACIONES

Dependencias y entidades involucradas	Departamento Nacional de Planeación <ul style="list-style-type: none">• Dirección de Economía Naranja y Desarrollo Digital (DENDD) - Unidad de Científicos de Datos• Dirección de Seguimiento y Evaluación de Políticas Públicas (DSEPP)
Encargado del proyecto	Diana Carolina Rodríguez (drodriguezcdnp.gov.co) Maria Camila Murillo (marimurillo@dn.gov.co)
Fecha de elaboración	25/07/2022
Versión	1.0

Contenido

1. Introducción	2
2. Metodología	2
3. Resultados	7
4. Observaciones y correcciones adicionales	20
5. Socialización	20



1. Introducción

En el marco del proyecto Medición del Uso de Evaluaciones siguiendo los lineamientos del plan de trabajo aprobado para su desarrollo, la UCD presenta este documento como el informe final que recopila los resultados del proyecto correspondientes a los tres entregables acordados.

En este documento se muestran los resultados encontrados del conteo de términos relacionados a la medición de las evaluaciones a políticas públicas en documentos CONPES en el periodo (2002-2021) y los PND entre (2002- 2022), los indicadores normalizados de estos conteos, para una mejor comparación entre documentos, un análisis de contexto, es decir el conteo de las palabras que se encuentran alrededor de los términos relacionados al uso de evaluaciones y el conteo de términos relacionados al uso de evaluaciones en Twitter. Para realizar este análisis se realiza un proceso de lectura, limpieza, adecuación y análisis de los documentos, en el proceso de análisis se realiza el conteo de los términos relacionados al uso y medición de evaluaciones planteados inicialmente y de los términos alrededor de estos, así como observar la frecuencia de términos para 36 cuentas en Twitter.

El documento se organiza de la siguiente manera, primero, se presenta la metodología usada para el análisis de texto, el proceso de lectura, limpieza, análisis de los documentos CONPES y PND, así como la metodología del análisis de contexto, del cálculo de los indicadores normalizados y el conteo de términos en Twitter seguido de los resultados de los documentos que se analizaron y las cuentas de Twitter que se revisaron para hacer el conteo. Finalmente, la Sección 4 contiene algunas observaciones que es importante tener en cuenta dados los resultados del proyecto.

2. Metodología

2.1. Lectura de los documentos

En esta primera etapa se realiza la lectura de todos los CONPES y PND disponibles. Para algunos de ellos se utiliza el OCR (Reconocimiento óptico de caracteres) que facilita la extracción de texto de imágenes o documentos escaneados. Una vez se realiza este procedimiento se obtienen unos documentos en formato .txt que permiten un posterior análisis.

2.2. Limpieza de los documentos

La etapa de limpieza de texto se aplica tanto para los documentos CONPES y PND, como a cada uno de los términos de interés que se buscarán en los documentos, este proceso que incluye 4 pasos:

1. Remover signos de puntuación y conversión a minúsculas: se retiran todos los signos de puntuación del texto tales como comas, puntos, signos de exclamación. Después todo el texto es convertido a texto en letras minúsculas.
2. Remoción de *stopwords* o palabras vacías: estas palabras no tienen significado por si solas, sino que modifican o acompañan a otras. Este conjunto se encuentra conformado por artículos, pronombres, preposiciones y adverbios.
3. Quitar palabras propias del contexto: Se remueve del texto palabras que pueden ser muy frecuentes dado la naturaleza del estudio, para ello se removieron nombres propios de personas, municipios y departamentos de Colombia y números.
4. Lematización de texto: se refiere a la transformación de todas las formas flexionadas de una palabra en su lema, por ejemplo, el plural, femenino y diminutivo de una palabra se transforma en su lema: mesas, mesón y mesitas se transforman todas en su lema, mesa.



Este proceso reduce el tamaño del vocabulario (número de palabras que conforman todo el universo de los documentos) lo que facilita la búsqueda automática y coincidencia de los términos de interés dentro de los documentos CONPES y PND.

2.3. Búsqueda y frecuencia de términos en los documentos

En este caso se implementa una estrategia de búsqueda con el fin de conocer el número de apariciones de los términos predefinidos en los documentos, así:

Dado a que un término o frase de interés se puede presentar de diferentes maneras en los documentos PDT, la estrategia de búsqueda permite hacer búsquedas flexibles de términos basado en expresiones regulares¹. Por ejemplo, la palabra *natural* aparece contenida en la palabra *naturaleza*, sin embargo, son dos palabras distintas que deben analizarse por separado, para dar solución a este problema se establece que la palabra *natural* sea contada solo si aparece idéntica, para ello la expresión regular es `\bnatural\b`.

Por otro lado, los términos compuestos por dos o más palabras, se buscan por separado en las ventanas de análisis, por ejemplo, la frase “Uso y aprovechamiento de datos”, es buscada como `uso|aprovechar|datos`, la raya vertical o pleca (|) actúa como una condición lógica “o” y permite que se busquen las cuatro expresiones al mismo tiempo. En la 1 se ilustra su funcionamiento sobre un texto de prueba.

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 1, hay 3 condiciones y cada una tiene 2 ocurrencias a excepción del término “datos” que tiene 4. Para encontrar el número de apariciones de la frase “Uso y aprovechamiento de datos”, se halla el mínimo de ocurrencias entre todas las condiciones, en este caso 2, por lo que la estrategia de búsqueda da como resultado del conteo 2 ocurrencias para esa ventana de análisis.

Figura 1: Ejemplo del funcionamiento de la búsqueda de términos



Fuente: Elaboración propia

Este proceso se hace para cada documento CONPES y PND. Al final se consolidan los conteos de los términos para todo el documento. Es importante tener en cuenta que algunas palabras pueden presentar conteos duplicados dado

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



que aparecen solas y también al interior de una frase, así como en el ejemplo anterior con el término “datos” en la frase datos abiertos y en la oración “aprovechamiento de los datos”.

Después de identificar todas las palabras que cumplen el patrón de la expresión regular, se hace un conteo por cada expresión de manera que el resultado se almacene en una matriz de datos, esto para elaborar los diagramas de frecuencia.

Los términos buscados en cada documento son:

- Evaluación de intervenciones/programas/políticas
- Resultados de evaluación de intervenciones/programas/políticas
- Toma de decisiones basada en evidencia
- Diseño de políticas basado en evidencia
- Uso de evaluaciones de intervenciones/programas/políticas
- Resultados de programa
- Resultados de política
- Efectos de programa/política
- Impactos de programa/política
- Evidencia de programa/política
- Evaluación institucional
- Evaluación de operaciones
- Evaluación de resultados
- Evaluación de impacto
- Análisis costo beneficio
- Análisis costo efectividad
- Evaluación ejecutiva
- Recomendación/recomendaciones de evaluaciones.

2.4. Normalización del conteo de palabras

Después de realizar el conteo de los términos de interés, se realizan dos normalizaciones, una por número de ventanas de cada documento (250 palabras por ventana), y segundo una normalización por año, en la que se divide la frecuencia de aparición de los términos relacionados al uso de evaluaciones en los CONPES o PND de un año en particular entre el número de documentos CONPES o PND analizados en dicho año.

El indicador normalizado por ventanas nos permite hacer comparaciones entre documentos, ya que cuantifica el número promedio de palabras que aparecen en cada ventana, siendo una ventana una unidad homogénea (agrupación de 250 palabras). Este indicador se interpreta como el número de palabras promedio relacionadas al uso de las evaluaciones que se encuentran por cada 250 palabras.

El indicador normalizado por el número de documentos CONPES/PND analizado, se interpreta como el promedio de palabras o frases relacionadas al uso de las evaluaciones en cada documento CONPES o PND, en cada año.

Indicador del conteo del término i normalizado por ventanas: Se divide el número de veces que aparece una palabra (frecuencia) o frase en los CONPES y PND entre el número de ventanas de dichos documentos:

$$\frac{\sum_{i \in j} \text{Frecuencia de palabra } i \text{ en los CONPES o PND}}{\text{Número de ventanas en los CONPES o PND}}$$



- **Indicador del conteo del término i normalizado por No. Documentos CONPES por año:** Se divide el número de veces que aparece una palabra (frecuencia) o frase en los CONPES y PND entre el número de CONPES o PND de cada año.

$$\frac{\sum_{i \in j} \text{Frecuencia de palabra } i \text{ en los CONPES o PND en el año } i}{\text{Número de CONPES o PND en el año } i}$$

2.5. Búsqueda de palabras contexto

Después de tener los resultados de la búsqueda de los términos relacionados a la Evaluación de Políticas Públicas se identifican las palabras que aparecen con mayor frecuencia alrededor de cada uno de los términos relacionados a la Evaluación de Políticas Públicas. El proceso se realiza buscando los unigramas/bigramas, que se encuentran dentro de las ventanas donde está contenido cada término, y contando la frecuencia de aparición de cada uno de ellos.

En la Figura 2 se observa un ejemplo de la búsqueda de términos relacionados con las palabras de interés.

Figura 2. Ejemplo conteo de bigramas relacionados con los términos TIC

Ventana 1, donde aparece el término tic “gobierno digital”

Se debe diferenciar entre la gobernanza digital y la institucionalidad que se hará cargo del **gobierno digital**. En tanto que la primera define los alcances y contenidos, la **política pública**, el marco normativo, el liderazgo, la infraestructura y las soluciones comunes (entre otros temas transversales), la segunda plantea el esquema de articulación organizacional (funciones, perfiles, competencias) e institucionalidad (ley, decreto o norma que le da atribuciones) encargado de brindar los servicios y las soluciones digitales para los servicios del Estado.

Ventana 2, donde aparece el término tic “gobierno digital”

El uso de las tecnologías digitales como parte integral de la **política pública** y las estrategias de modernización de los gobiernos con el fin de crear valor público. Esto se basa en un ecosistema de **gobierno digital** constituido por los actores estatales, organizaciones no gubernamentales, empresas, asociaciones de ciudadanos y personas encargadas de la producción y acceso a los datos, servicios y contenidos a través de interacciones con el gobierno

Fuente: Elaboración Propia

Para el caso del ejemplo, el término de interés es “*gobierno digital*”, y se encuentra el bigrama (grupo de dos palabras) “*política pública*” una vez en la ventana 1 donde aparece por primera vez el término TIC y otra vez en la ventana 2 donde aparece por segunda vez el término TIC. De manera que el término “*política pública*”, es un bigrama de contexto relacionado al término TIC “*gobierno digital*” que tiene una frecuencia de dos.



2.6. Conteo de Términos en Twitter

2.6.1. Recolección de Tweets:

Para esta etapa se descargaron todos los tweets asociados a las 36 cuentas de interés (ver Tabla 1). El análisis se realizó para el periodo de tiempo comprendido entre el 1 de enero de 2014 y el 31 de mayo de 2022 pues antes de este periodo hay inconvenientes para encontrar información sobre los términos relacionados al uso de evaluaciones. Este proceso se realizó con Python a través de la librería Tweepy.

Tabla 1. Cuentas de Interés

Rf	Entidad	Twitter 1	Twitter 2	Twitter 3
1	MINISTERIO DE CULTURA	@mincultura		
2	MINISTERIO DE TRANSPORTE	@MintransporteCo		
3	MINISTERIO DE TECNOLOGÍAS DE LA INFORMACIÓN Y LAS COMUNICACIONES	@Ministerio_TIC		
4	MINISTERIO DE VIVIENDA, CIUDAD Y TERRITORIO	@Minvivienda		
5	MINISTERIO DE AMBIENTE Y DESARROLLO SOSTENIBLE	@MinAmbienteCo		
6	MINISTERIO DE EDUCACIÓN NACIONAL	@Mineducacion		
7	MINISTERIO DE COMERCIO, INDUSTRIA Y TURISMO	@MincomercioCo		
8	MINISTERIO DE MINAS Y ENERGÍA	@MinEnergiaCo		
9	MINISTERIO DE TRABAJO	@MintrabajoCol		
10	MINISTERIO DE SALUD Y PROTECCIÓN SOCIAL	@MinSaludCol		
11	MINISTERIO DE AGRICULTURA Y DESARROLLO RURAL	Minagricultura		
12	MINISTERIO DE DEFENSA NACIONAL	@mindefensa		
13	MINISTERIO DE JUSTICIA Y DERECHO	@minjusticiaco		
14	MINISTERIO DE HACIENDA Y CRÉDITO PÚBLICO	@MinHacienda		
15	MINISTERIO DE RELACIONES EXTERIORES	@CancilleriaCol		
16	MINISTERIO DEL INTERIOR	@Mininterior		
17	MINISTERIO DE CIENCIAS, TECNOLOGIA E INNOVACIÓN	@MincienciasCo		
18	VICEPRESIDENCIA DE LA REPÚBLICA	@ViceColombia		
19	PRESIDENCIA	@infopresidencia		
20	PROSPERIDAD SOCIAL	@ProsperidadCol		
21	DEPARTAMENTO NACIONAL DE PLANEACIÓN	@DNP_Colombia		
22	CONSEJO PRIVADO DE COMPETITIVIDAD	@ColombiaCompite		
23	FEDESARROLLO	@Fedesarrollo		
24	ECONOMETRIA CONSULTORES	@econometria_sa		
25	CENTRO NACIONAL DE CONSULTORIA	@CNCSocial		
26	UNIVERSIDAD DE LOS ANDES	@Uniandes	@EconomiaUAndes	
27	UNIVERSIDAD DEL ROSARIO	@URosario	@EconUrosario	@Labo_UR
28	UNIVERSIDAD JAVERIANA	@UniJaveriana	@ofiscalpuj	@JaverianaEcon
29	EAFIT	@FEG_EAFIT	@EAFIT	
30	ECONOMÍA URBANA	@EconomiaUrbana_		

Fuente: Elaboración Propia.



2.6.2 Limpieza y adecuación de los Tweets

Una vez son descargados los tweets para las cuentas de interés se realiza el proceso de limpieza y adecuación, esto incluye remoción de caracteres especiales, limpieza de stopwords (palabras cuyo significado por sí solas es nulo, como preposiciones, conectores o artículos, las cuales no aportan valor al análisis). Una vez se realiza la limpieza el paso a seguir es la lematización de los textos, este es un proceso lingüístico que transforma las palabras de su forma flexionada (plural, femenino, conjugaciones, etc.) a su lema correspondiente que representa todas las formas flexionadas de una misma palabra. Por ejemplo, las palabras niños, niña y niño tienen el mismo lema que es niño. Este proceso contribuye a la simplificación de los textos evitando tener un vocabulario más grande del necesario. Con los Tweets ya adecuados se unifican todas las respuestas asociadas a los agentes de interés como resultado queda una base de datos de 9142 Tweets para realizar el proceso de búsqueda.

2.6.3. Búsqueda de palabras relacionadas al uso de evaluaciones

Con los Tweets encontrados en la etapa anterior, se realiza la búsqueda de las palabras relacionadas al uso de evaluaciones, esto se realiza a través de tablas de frecuencia que muestran para cada Tweet el número de veces que aparece cada uno de los términos relacionados al uso de evaluaciones.

De este proceso la base de datos se reduce a 82 Tweets que fueron publicados por las cuentas de interés y que contienen 118 términos relacionados al uso de evaluaciones. Para el conteo, se busca el cada una de las palabras del término y se cuenta si el tweet tiene las palabras que componen el termino sin importar el orden, de manera que un mismo tweet puede tener varias frases de interés. Por ejemplo el tweet *“tip del día **evaluacion impacto mejores políticas** publicas ejemplo foco económico”* de la cuenta de *“econometria_sa”*, se cuenta para los términos *“evaluación de políticas”* e *“impacto de políticas”*.

3. Resultados

3.1. Documentos inspeccionados

CONPES 2002-2021/PND 2002-2022

Se analizaron cinco (5) Planes Nacionales de Desarrollo-PND comprendidos entre el periodo de 2002 a 2022, que corresponden a seis documentos, ya que el PND 2010-2014 consta de dos tomos. En la tabla 2 se muestra el número de documentos PND analizados, en todos ellos se encuentran términos relacionados al uso de las evaluaciones.

Tabla 2. Documentos en estudio PND (2002-2022)

Año	No. Documentos PND analizados	Documentos con Términos relacionados al uso de evaluaciones	% Documentos con Palabras
2002-2006	1	1	100%
2006-2010	1	1	100%
2010-2014	2	2	100%
2014-2018	1	1	100%
2018-2022	1	1	100%

Fuente: Elaboración propia



Se examinaron 848 documentos CONPES entre el periodo (2002- 2021), de los cuales 336 documentos tenían al menos una palabra relacionada al uso de las evaluaciones, lo que corresponde al 40% de los documentos. En la tabla 3, se muestra el total de documentos CONPES analizados por año y el total de documentos que contienen al menos 1 término o frase relacionado al uso de las evaluaciones. Se observa un crecimiento en los documentos que contienen estos términos en los últimos años.

En la figura 3 se muestra el total de documentos CONPES analizados vs el número de documentos con palabras o frases relacionadas al uso de las evaluaciones.

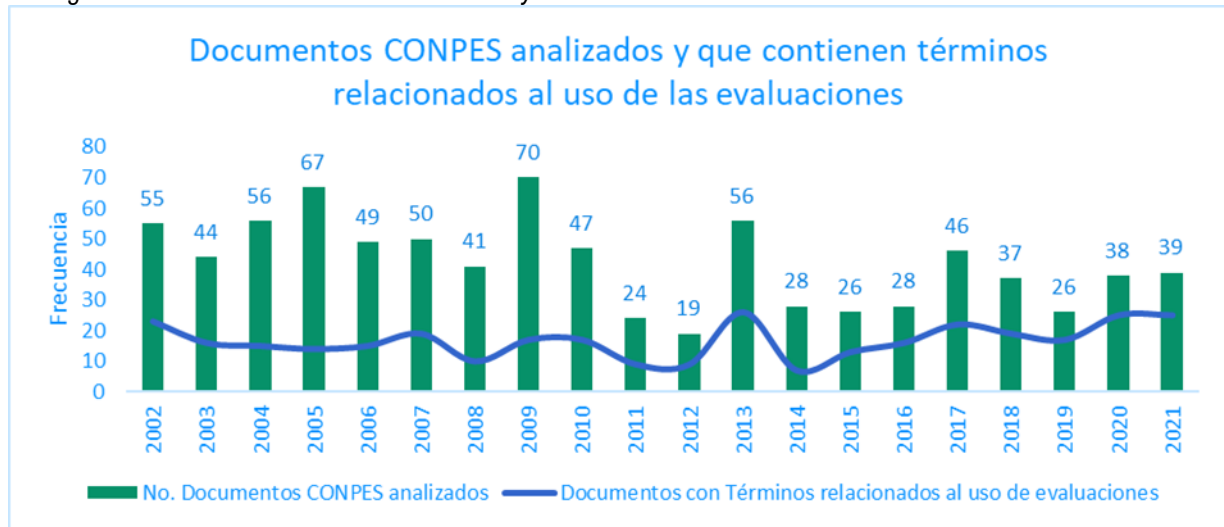
Tabla 3. Documentos en estudio CONPES (2002-2021)

Año	No. Documentos CONPES analizados	Documentos con Términos relacionados al uso de evaluaciones	% Documentos con Palabras
2002	55	23	42%
2003	44	16	36%
2004	56	15	27%
2005	67	14	21%
2006	49	15	31%
2007	50	19	38%
2008	41	10	24%
2009	70	17	24%
2010	47	17	36%
2011	24	9	38%
2012	19	9	47%
2013	56	26	46%
2014	28	7	25%
2015	26	13	50%
2016	28	16	57%
2017	46	22	48%
2018	37	19	51%
2019	26	17	65%
2020	38	25	66%
2021	39	25	64%
Total	848	336	40%

Fuente: Elaboración propia



Figura 3. Documentos CONPES analizados y documentos con términos relacionados al uso de evaluaciones



Fuente: Elaboración propia

A través del desarrollo metodológico descrito, se obtuvieron los resultados que se presentan a continuación. Toda retroalimentación desde un punto de vista experto o de usuario por parte de la DSEPP es bienvenida. Este insumo será de gran ayuda para mejorar la calidad y utilidad de los resultados obtenidos, de manera que agreguen mayor valor.

3.2. Resultado de los conteos:

Inicialmente, se verifica la evolución del número de términos relacionados al uso de las evaluaciones en los documentos CONPES, el año 2002 tiene el mayor número de apariciones con 198 términos, sin embargo, se evidencia una tendencia creciente en el número de términos relacionados al uso de las evaluaciones en los últimos años.

Figura 4. Conteo histórico de los términos relacionados al uso de las evaluaciones en los CONPES 2002-2021

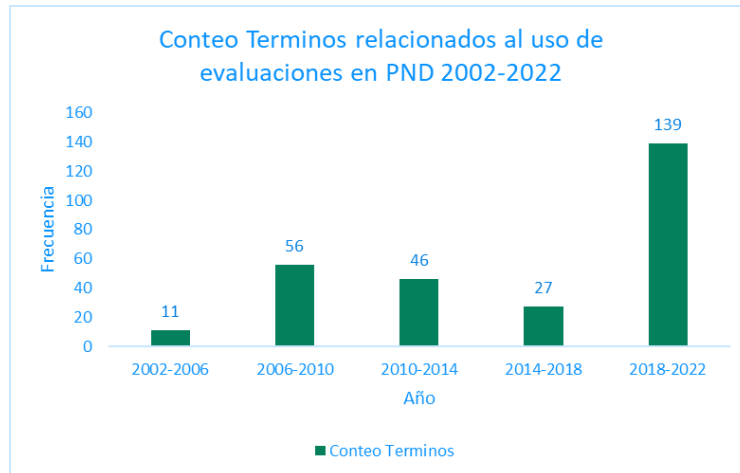


Fuente: Elaboración propia



Por otro lado, los documentos PND creados en el periodo (2018-2022) tienen más conteos por año con 139 términos a diferencia de los PND creados en (2002-2006) donde se hallaron simplemente 11 términos.

Figura 5. Conteo histórico de los términos relacionados al uso de las evaluaciones en los PND 2002-2022



Fuente: Elaboración propia

Paso seguido, observamos la frecuencia de los términos relacionados al uso de las evaluaciones, en los documentos CONPES 2002-2021, los términos de mayor frecuencia son “Evaluación de intervenciones/programas/política”, “Evaluación de impacto” y “Evaluación de resultados” con 407, 303 y 175 términos respectivamente. Los términos “Uso de evaluaciones de intervenciones/programas/políticas” y “Diseño de políticas basado en evidencia” no aparecieron registrados en ninguno de los documentos CONPES analizados.

Figura 6. Términos hallados por palabra CONPES (2002-2021)

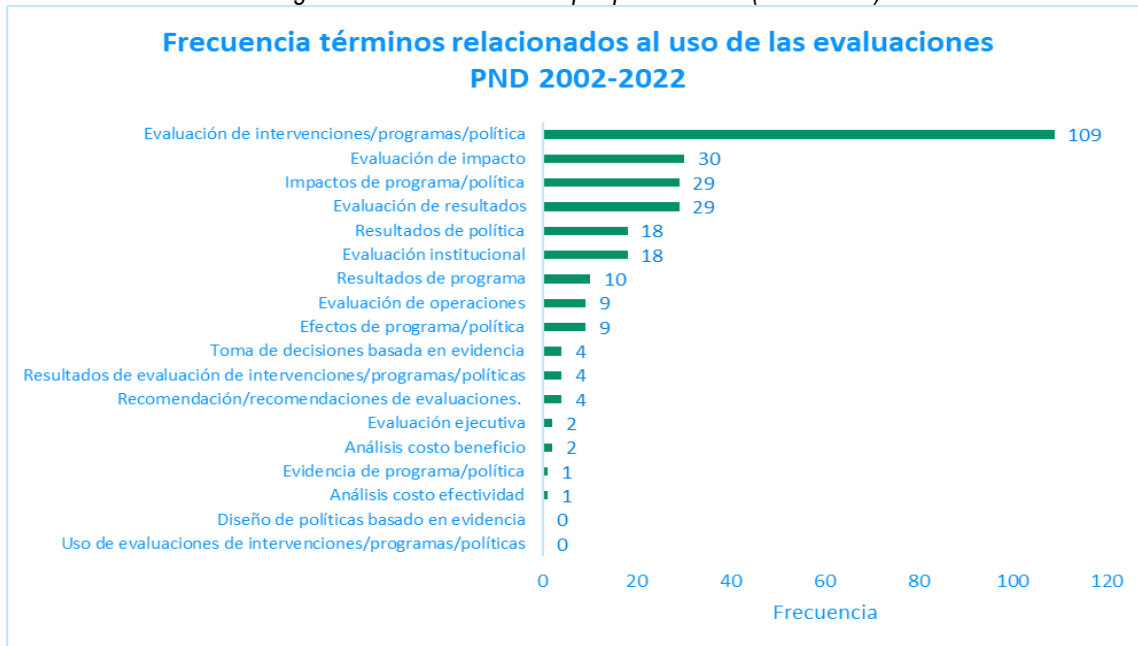


Fuente: Elaboración propia



Por otro lado, en la *figura 7* se observa que los términos “Evaluación de intervenciones/programas/política” y “Evaluación de impacto” tienen mayor frecuencia en los PND (2002-2022) con 109 y 30 términos respectivamente, en este punto es de considerar que se encontraron 16 términos de los 18 solicitados, ya que los términos “Uso de evaluaciones de intervenciones/programas/políticas” y “diseño de políticas basado en evidencia” no se encuentran en los documentos.

Figura 7. Términos hallados por palabra PND(2002-2022)



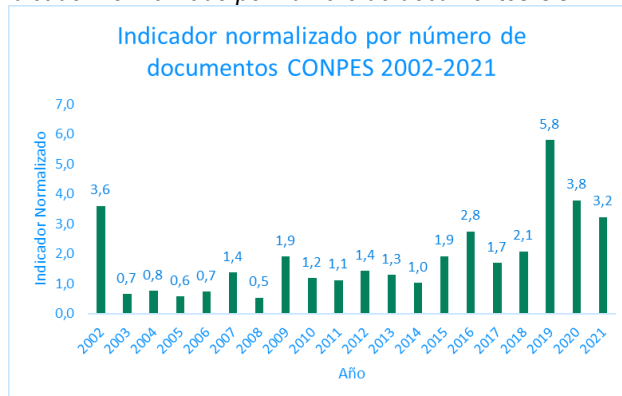
Fuente: Elaboración propia

3.3. Indicadores Normalizados

Índice normalizado por número de documentos CONPES (2002-2021) y PND (2002-2022):

Con el fin de tener un dato agregado que nos sintetice la información, se realizan una normalización de los datos, dividiendo el total de términos relacionados al uso de las evaluaciones encontrados entre el número de documentos CONPES analizados en cada año de estudio. Este índice nos indica el promedio de términos relacionados al uso de las evaluaciones en cada año.

Figura 8. Indicador normalizado por número de documentos CONPES 2002-2021



Fuente: Elaboración propia



Tabla 4. Conteo de términos relacionado al uso de las evaluaciones, número de documentos CONPES analizados e indicador normalizado por año

Año	Conteo Terminos	No. Documentos CONPES	Ind.Normalizado
2002	198	55	3,6
2003	29	44	0,7
2004	43	56	0,8
2005	39	67	0,6
2006	36	49	0,7
2007	69	50	1,4
2008	22	41	0,5
2009	134	70	1,9
2010	56	47	1,2
2011	27	24	1,1
2012	27	19	1,4
2013	73	56	1,3
2014	29	28	1,0
2015	50	26	1,9
2016	77	28	2,8
2017	78	46	1,7
2018	77	37	2,1
2019	151	26	5,8
2020	144	38	3,8
2021	126	39	3,2
Total	1485	846	1,8

Fuente: Elaboración propia

Tabla 5. Indicador Normalizado PND (2002-2022)

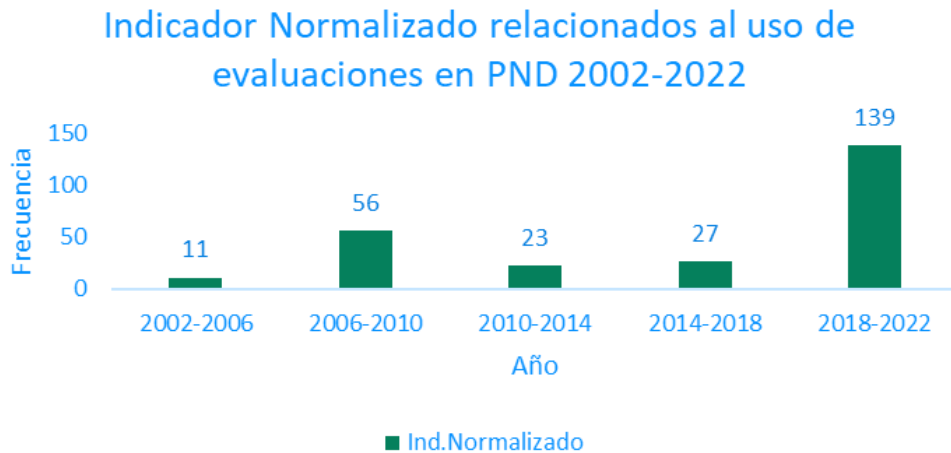
Año	Conteo Terminos	No. Documentos PND	Ind.Normalizado
2002-2006	11	1	11
2006-2010	56	1	56
2010-2014	46	2	23
2014-2018	27	1	27
2018-2022	139	1	139
Total	279	6	47

Fuente: Elaboración propia



Cuando se hace la normalización del conteo de términos para los documentos PND(2002/2022) , el indicador y el conteo es el mismo, ya que hay un documento PND por periodo excepto para el periodo 2010-2014, cuando se tienen dos tomos.

Figura 9. Gráfico de barras del conteo de términos normalizado para los PND(2002-2022)

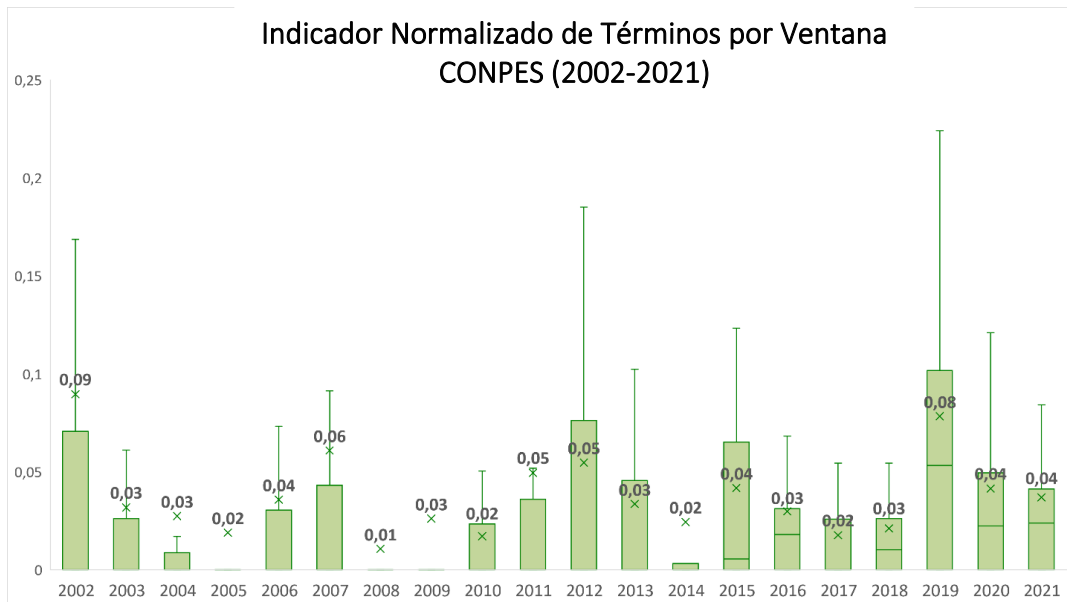


Fuente: Elaboración propia

Índice normalizado por número de ventanas CONPES (2002-2021) y PND (2002-2022):

También se realiza la normalización por el número de ventanas con el fin de tener un parámetro estándar de comparación, ya que cada documento CONPES o PND tiene un número de páginas diferente y las páginas a su vez tienen un número de palabras diferentes. Siendo una ventana una agrupación de 250 palabras, lo que nos señala el indicador es el número de términos promedio relacionados al uso de las evaluaciones por cada 250 palabras.

Figura 10. Indicador normalizado por ventanas CONPES (2002-2021)



Fuente: Elaboración propia



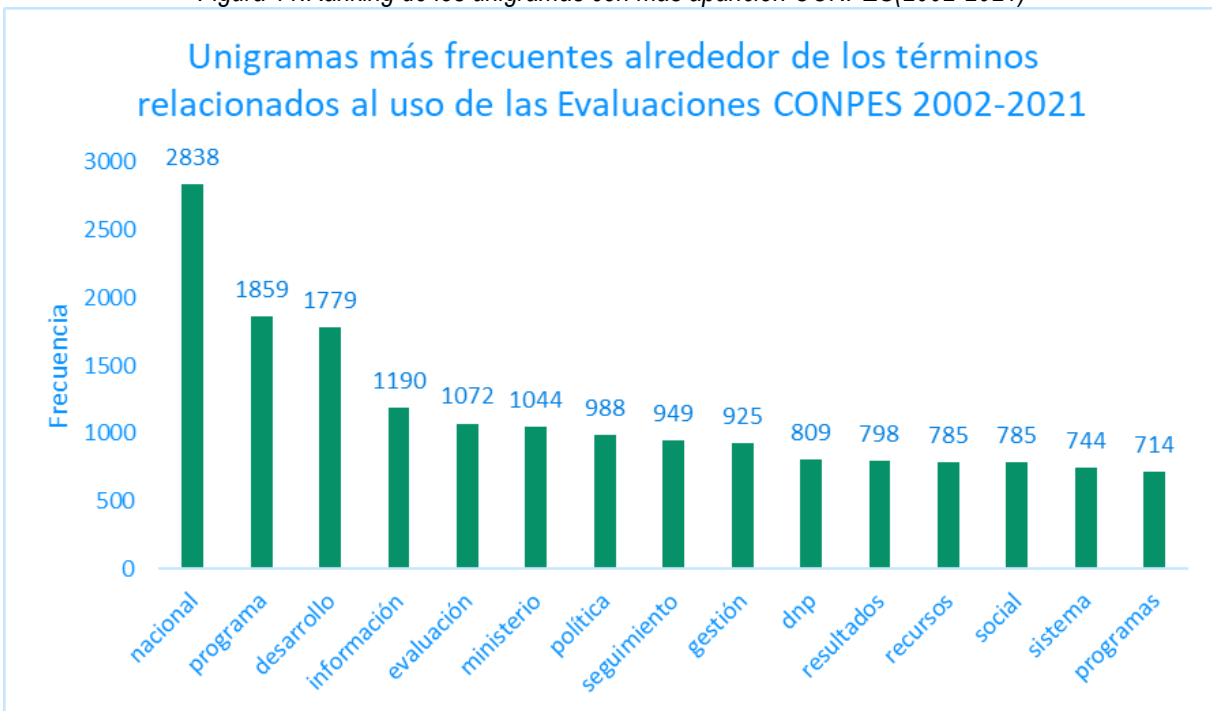
Para el cálculo del indicador, se divide el número de términos encontrados en cada documento CONPES/PND en el total de ventanas del documento (número de palabras dividido en 250). Para representar los resultados de los indicadores, en la figura 10 se realiza un diagrama de cajas que muestra la variabilidad de los indicadores en los diferentes documentos CONPES por cada año, el promedio de los indicadores en cada año se señala con una cruz y el máximo con la terminación de la línea. Los documentos CONPES desarrollados en el año 2002 y 2019 son los que tienen el mayor índice promedio, con un valor de 0.09 y 0.08 respectivamente, mientras que los documentos realizados en el año 2008 tienen el menor valor en el indicador con 0.01. Así mismo, se puede observar en el Anexo 2 el indicador normalizado detallado por término, por documento y por año.

El indicador nos señala, el promedio de términos relacionados al uso de las evaluaciones en cada 250 palabras, por ejemplo en el año 2021 el indicador con un valor de 0,04 nos indica que en promedio se encuentran 0,04 términos relacionados al uso de las evaluaciones por cada 250 palabras, o lo que es lo mismo, por cada 25000 palabras se encuentran 4 términos relacionado al uso de las evaluaciones.

3.4. Análisis de Contexto

El paso siguiente es encontrar las palabras que acompañan los 18 términos predefinidos para buscar en los documentos CONPES y PND, como resultado los unigramas más frecuentes que aparecen en los documentos CONPES(2002-2021) en terminos agregados corresponden a nacional, programa, desarrollo e información.

Figura 11. Ranking de los unigramas con más aparición CONPES(2002-2021)

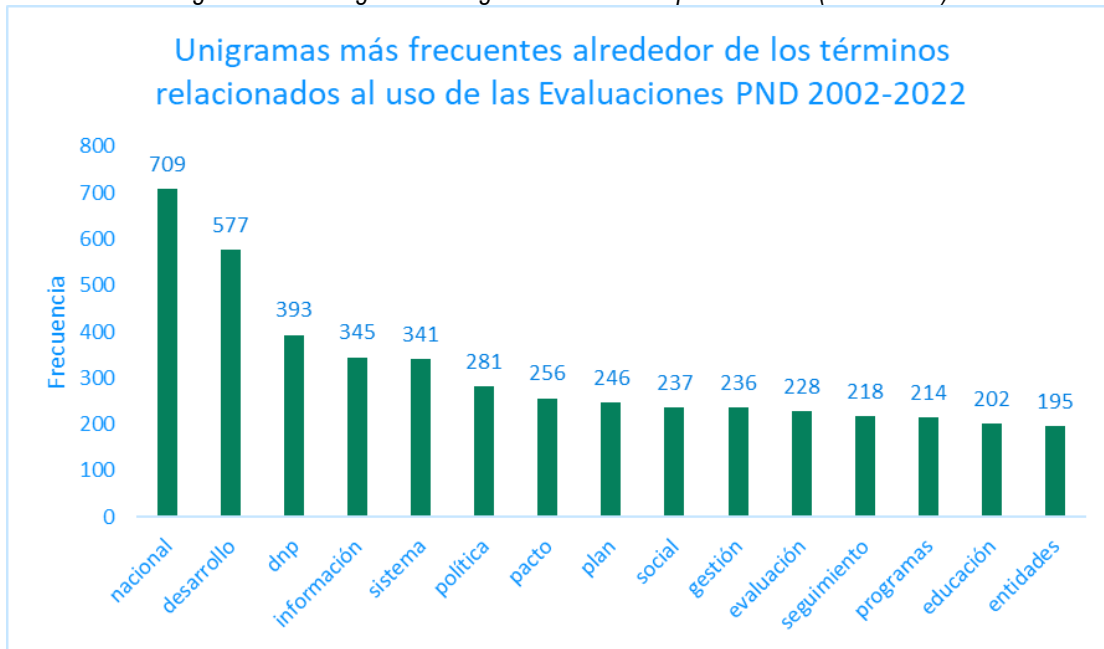


Fuente: Elaboración propia

En los documentos PND(2002-2022) los unigramas más frecuentes que acompañan los términos de interés definidos son nacional, desarrollo y dnp (figura 12)



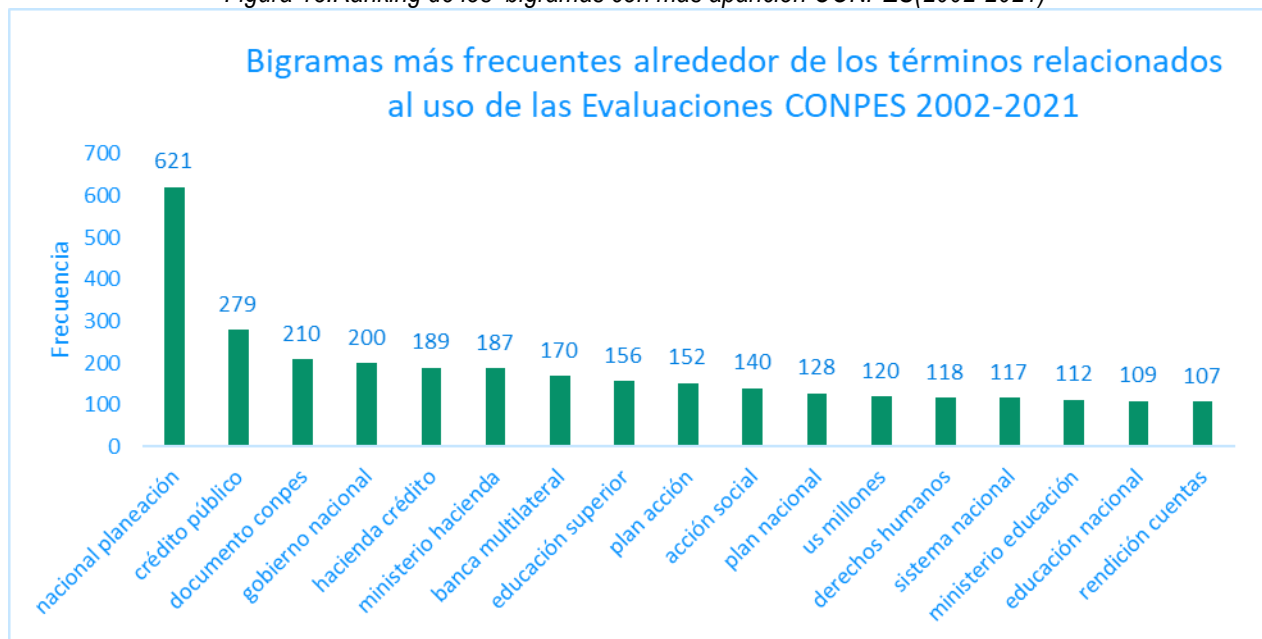
Figura 12. Ranking de los unigramas con más aparición PND(2002-2022)



Fuente: elaboración propia

Después se encontraron los bigramas (conjunto de dos palabras) que acompañan los términos relacionados al uso de las evaluaciones, entre los bigramas más frecuentes en términos agregados se encuentran (nacional planeación, crédito público, documento conpes y gobierno nacional) en los documentos CONPES(2002-2021) (Figura 13). Mientras que en los documentos PND(2002-2022) los bigramas de contexto más frecuentes son (plan nacional, nacional desarrollo y pacto equidad) (Figura 14). De igual manera en el Anexo 3 se encuentra la frecuencia de bigramas por término.

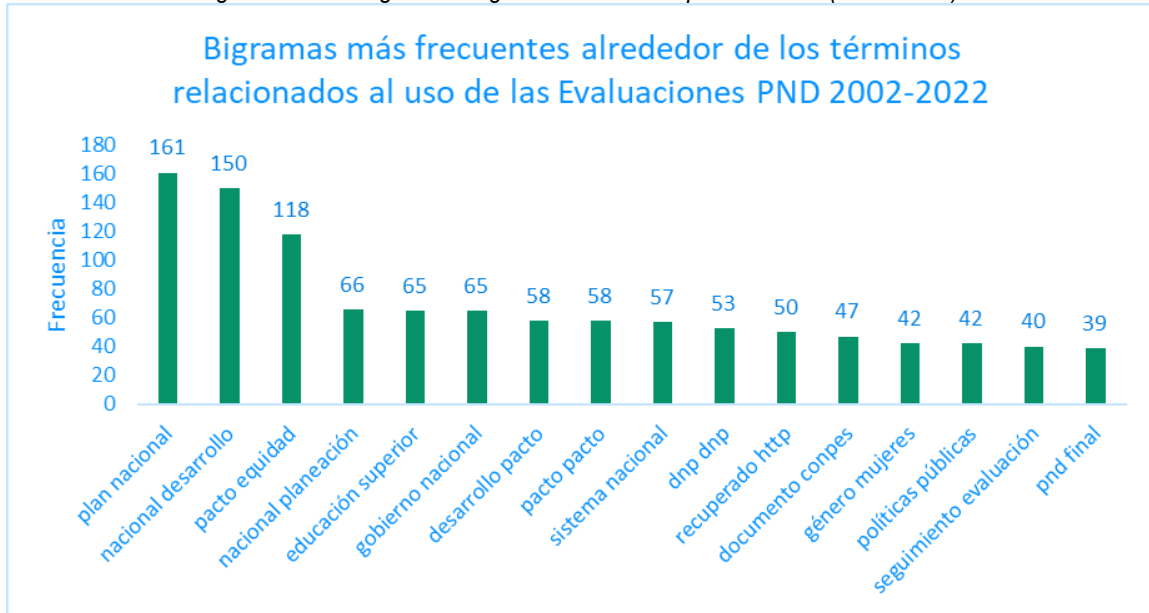
Figura 13. Ranking de los bigramas con más aparición CONPES(2002-2021)



Fuente: elaboración propia



Figura 14. Ranking de los bigramas con más aparición PND(2002-2022)

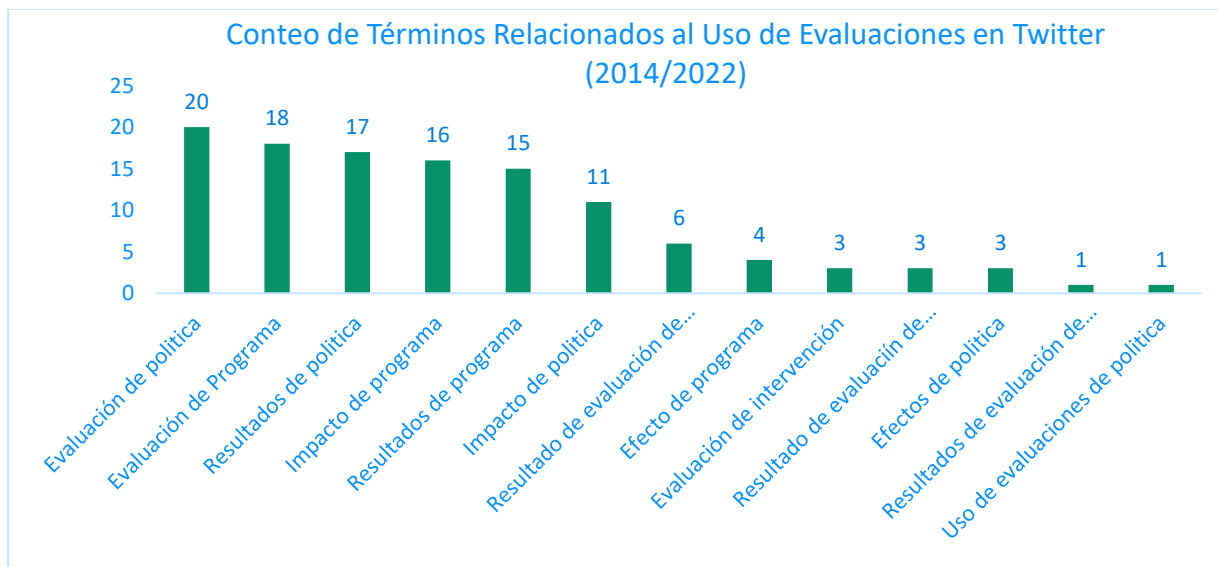


Fuente: Elaboración propia

3.5 Conteo de términos en Twitter

En este punto se realiza la descarga de los tweets con los términos relacionados al uso de las evaluaciones y los agentes de interés, se realiza el conteo de cada uno de estos términos a partir del año 2014 hasta el 2022. Como resultado se obtiene que los términos con más conteos son evaluación de política, evaluación de programa y resultados de política con 20, 18 y 17 apariciones respectivamente, a diferencia de los términos con menos conteos como Uso de Evaluación de política y resultado de evaluación con 1 aparición cada una.

Figura 15. Conteo de términos relacionados al uso de evaluaciones (2014-2022)



Fuente: elaboración propia



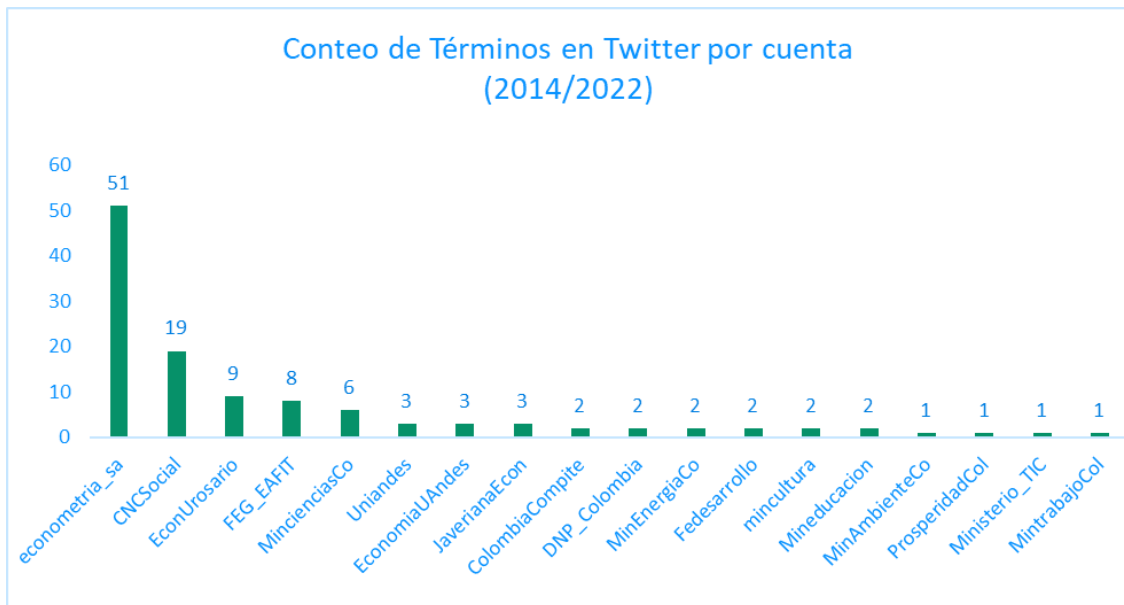
Por otro lado, en el periodo (2014/2022) se encuentran más apariciones de los términos relacionados al uso de evaluaciones en la cuentas econométrica_sa:51 apariciones, CNCSocial:19 apariciones y EconUrosario:9 apariciones, a diferencia de las cuentas ubicadas en las últimas posiciones que solo tienen una aparición relacionada a los términos de interés. (tabla 6).

Tabla 6. Términos encontrados por cuenta (2014-2022)

Cuenta	Términos relacionados al uso de evaluaciones
econométrica_sa	51
CNCSocial	19
EconUrosario	9
FEG_EAFIT	8
MincienciasCo	6
Uniandes	3
EconomíaUAndes	3
JaverianaEcon	3
ColombiaCompite	2
DNP_Colombia	2
MinEnergíaCo	2
Fedesarrollo	2
mincultura	2
Mineducación	2
MinAmbienteCo	1
ProsperidadCol	1
Ministerio_TIC	1
MintrabajoCol	1
Total	118

Fuente: Elaboración propia

Figura 16. Conteo de términos relacionados al uso de evaluaciones por cuenta (2014-2022)

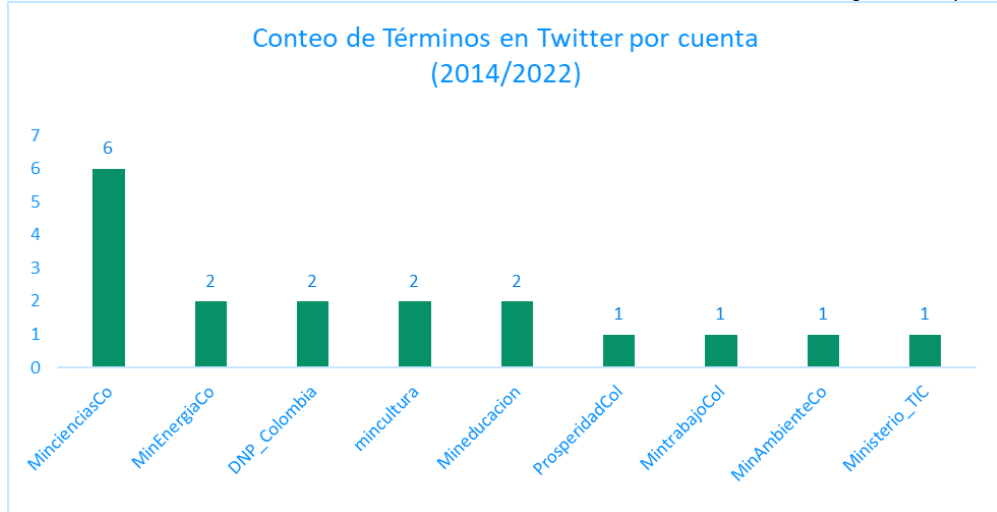


Fuente: Elaboración propia



Al analizar los términos encontrados en las cuentas de Twitter del gobierno, los términos relacionados al uso de evaluaciones que más apariciones tiene es MincienciasCo: 6 apariciones, a diferencia de las cuentas ProsperidadCol, MintrabajoCol, MinAmbienteCo y Ministerio_Tic con una aparición.

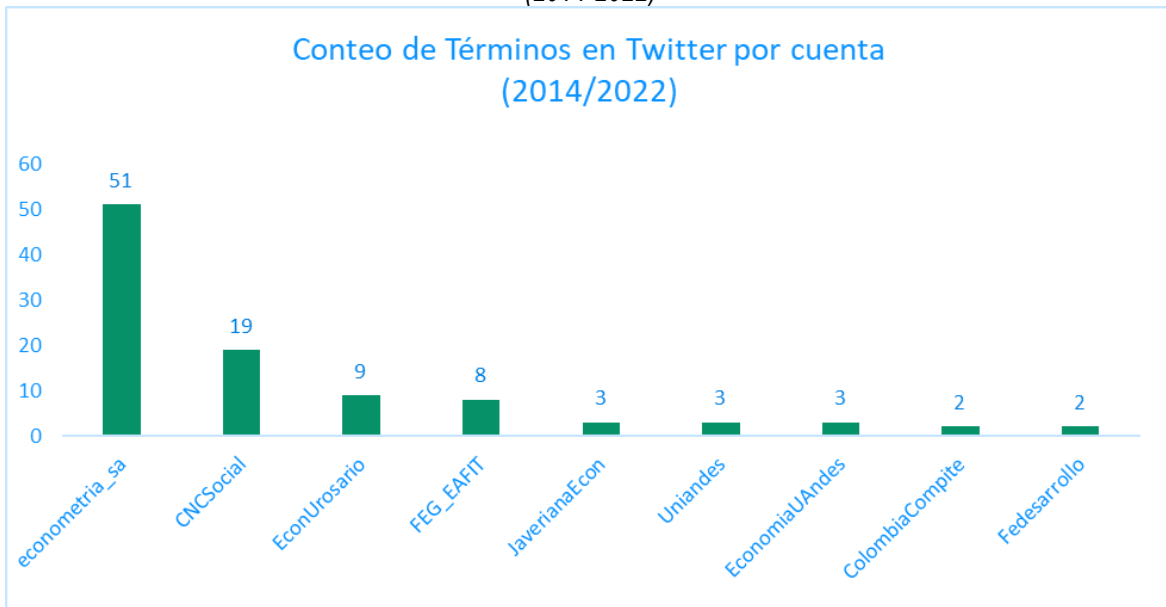
Figura 17. Conteo de términos relacionados al uso de evaluaciones en las cuentas del gobierno (2014-2022)



Fuente: Elaboración propia

Por otro lado, se evidencia mayor participación de los términos relacionados al uso de evaluaciones en los Tweets de las cuentas de consultorias y universidades en comparación a las cuentas del gobierno, así pues las cuentas con más términos encontrados son :econométrica_sa: 51 apariciones, CNCSocial:19 apariciones y las cuentas con menos términos encontrados son Fedesarrollo y ColombiaCompite con 2 apariciones cada una.

Figura 18. Conteo de términos relacionados al uso de evaluaciones en las cuentas de consultoria y universidades (2014-2022)



Fuente: Elaboración propia



Los términos relacionados al uso de evaluaciones en su mayoría fueron encontrados de manera separada, así:

- ✓ Término: “Resultados de Política”
- ✓ Cuenta: @CNCSocial
- ✓ Publicación (Texto Preprocesado):

“conozca reciente encuesta cncsocial revista semana percepciones expectativas ciudadanas realidad social política colombiana ambiente preelectoral aqui resultados recomendado”

Se tienen los siguientes anexos:

1. En el anexo 1 “*Conteo_términos_relacionados_uso_evaluaciones*” se encuentra el detalle del conteo de cada uno de los términos en los documentos analizados, tiene las columnas año, documento, tipo de documento y una columna por cada término.

Año	Documento	Tipo de Documento	Término 1	Término 2	...	Término n	Total Terminos
2002	3156	CONPES	0	0	0	0	0
2002	3157	CONPES	0	0	0	0	0
2002	3159	CONPES	0	0	0	0	1

2. En el Anexo 2 “*Indicadores normalizados*” se encuentra una pestaña con los conteos, el número de páginas y de ventanas por documento y en la otra pestaña, los indicadores normalizados detallados para cada uno de los términos de estudio.

Año	Documento	Tipo de Documento	Ventanas de 250 palabras	Indicador Total Normalizado	Indicador normalizado	Indicador normalizado	...	Indicador normalizado
2002	3156	CONPES	5,9	0	0	0	0	0
2002	3157	CONPES	7,8	0	0	0	0	0
2002	3159	CONPES	5,6	0,2	0	0	0	0

3. En el Anexo 3 “*Conteo_terminos_contexto*” se encuentra el conteo detallado de los unigramas y bigramas, que se encontraron alrededor de cada término de interés en cada uno de los documentos analizados, tiene la siguiente estructura, con una pestaña para unigramas y otra para bigramas.

Año	Documento	Tipo Documento	Palabra	Unigrama	Conteo
2002	3159	Conpes	Impactos de programa/política	recursos	6
2002	3159	Conpes	Impactos de programa/política	social	5
2002	3159	Conpes	Impactos de programa/política	condiciones	4

4. En el Anexo 4 “*Conteo_Terminos_Twitter*” se encuentra el conteo detallado de los términos relacionados al uso de evaluaciones encontrados en las cuentas de interés acompañado de una pestaña con los gráficos , la estructura es la siguiente:



Año	Cuenta	Texto	Termino 1	...	Termino n	Total
2014	CNCSocial	encuesta cnc resultados	0		0	1

4. Observaciones y correcciones adicionales

- De los términos de interés, los de mayor número de apariciones son “*evaluación de intervenciones/programas/políticas*”, “*evaluación de impacto*” y “*evaluación de resultados*”, mientras que para los Planes Nacionales de Desarrollo son “*evaluación de intervenciones/programas/políticas*”, “*evaluación de resultados*” e “*impactos de programa/política*”.
- El indicador normalizado por ventanas, tiene la finalidad de comparar la inclusión de los términos de interés en textos de diferentes características (tamaño de letra, espacios, número de gráficas y tablas, etc), ya que nos indica el número promedio de términos relacionados al uso de las evaluaciones por cada 250 palabras. Mientras que el indicador normalizado por número de documentos, nos indica el promedio de términos en cada CONPES o PND.
- El análisis de contexto nos permite conocer las palabras que se encuentran alrededor de los términos relacionados al uso de las evaluaciones, los unigramas más frecuentes en los documentos CONPES son “*nacional*”, “*desarrollo*” y “*programa*”, y en los PND son “*nacional*”, “*seguimiento*” e “*información*”.
- Al realizar el proceso de conteo de términos en Twitter resulta que la cantidad de términos relacionados al uso de evaluaciones es baja teniendo en cuenta el periodo (2014/2022), lo que puede ser consecuencia de las restricciones de la aplicación para la extracción de tweets.

5. Socialización

Los resultados de este proyecto se han socializado con la Dirección de Seguimiento y Evaluación de Políticas Públicas (DSEPP) durante el desarrollo de este y una vez se contaron con todos los entregables.

Contacto

Si tiene alguna duda, comentario o sugerencia sobre este proyecto, o si le gustaría conversar con la Unidad de Científicos de Datos sobre la posibilidad de una nueva fase para el mismo, puede comunicarse con nosotros a través del correo electrónico ucd@dnpp.gov.co.