

Analítica aplicada al Plan Nacional de Desarrollo 2022-2026: “Colombia, potencia mundial de la vida”

Enrique Arrieta-Díaz¹
Universidad del Magdalena
earrieta@unimagdalena.edu.co

Rafael Maldonado-Agudelo²
Universidad del Magdalena
rafaelmaldonadoaa@unimagdalena.edu.co

Amanda Vargas-Prieto³
Universidad del Magdalena
avargasp@unimagdalena.edu.co

DOI:

Fecha de recepción: 13 de febrero de 2024

Fecha de aprobación: 25 de abril de 2024



Cómo citar este artículo: Arrieta-Díaz, E.; Maldonado-Agudelo, R.; Vargas-Prieto, A. (2024). Analítica aplicada al Plan Nacional de Desarrollo 2022-2026: “Colombia, potencia mundial de la vida”. *Revista Escuela de Administración de Negocios*, (96), (páginas). DOI:

Resumen

Teniendo en cuenta la importancia de los planes de desarrollo para un país, y las dificultades de entender las propuestas de manera objetiva y neutra. Este artículo propone el análisis del PND 2022-2026: “Colombia, potencia mundial de la vida”, a través de la minería de texto. Los principales resultados fueron: la identificación de los 23 temas dentro del PND y los términos que incluyen. El término más importante es Colombia, seguido de paz, discapacidad, género y alimentos. Los dos primeros representan los temas con el mismo nombre. Siendo coherente con el eje transversal del PND, la consecución de la paz total en Colombia. Los otros tres términos están dentro de los temas de mujeres e inclusión, alimentación y mujeres, nuevamente. El análisis permitió la identificación de los bigramas y trigramas con mayor frecuencia en el PND. Siendo el más frecuente gobierno-nacional, mientras que el trigramas más frecuente es: plan-nacional-desarrollo. Este análisis evidencia que en el PND, la paz, las minorías, la alimentación, la educación, el transporte y los derechos humanos son protagonistas. La inversión es de los temas más relevantes en el PND. Los dos aspectos centrales del tema educación y salud se encuentran en una variedad de aspectos que hacen que su especificidad sea baja y su relevancia sea media. Sin embargo, es un tema transversal en el PND. Al revisar la correlación de los temas se observa que los temas fundamentales del PND son: infraestructura, paz, agua, desarrollo rural y cultura y recreación.

Palabras clave: analítica de datos; planeación participativa; administración pública; políticas públicas; minería de texto; Colombia.

¹Ph.D. Física. Michigan State University. M.Sc. y B.Sc. física. Universidad de los Andes. Investigador Asociado. Quantil. Profesor. Universidad del Magdalena, Santa Marta, Colombia. ORCID: <https://orcid.org/0000-0002-2408-3379>

²Estudiante de Ingeniería de Sistemas. Universidad del Magdalena, Santa Marta. ORCID: <https://orcid.org/0009-0001-5133-1040>

³ Ph.D. Ciencias Económicas. Universidad de Burdeos. Francia. Magister en Inteligencia Económica y Estrategias Competitivas. Universidad de Angers. Francia. Administradora de Empresas. Consejera Nacional de Planeación, Representante del sector educativo y cultural. Profesora. Universidad del Magdalena, Santa Marta, Colombia. ORCID: <https://orcid.org/0000-0002-8508-6979>

Analytics applied to the National Development Plan 2022-2026: "Colombia, world power of life"

Abstract

Considering the importance of Development Plans for a country, and the difficulties of understanding the proposals in an objective and neutral way. This article proposes the analysis of the PND 2022-2026 Colombia, world power of life through text mining. The main results were the identification of the 23 themes within the PND and the terms that include these themes. The most important term is Colombia, followed by peace, disability, gender, and food. The first two terms represent the themes with the same name. Being coherent with the transversal axis of the PND, the achievement of total peace in Colombia. The other three terms are within the themes of women and inclusion, food, and women again. The analysis allowed the identification of the bi-grams and tri-grams with greater frequency in the PND. Being the most frequent national-government, while the most frequent trigram is plan-national-development. This analysis shows that in the PND, peace, minorities, food, education, transportation, and human rights are protagonists. Investment is one of the most relevant issues in the PND. The two central aspects of the Education and Health theme are found in a variety of aspects that make their specificity low and their relevance medium. However, it is a cross-cutting issue in the PND. When reviewing the correlation of the themes, it is observed that the fundamental themes of the PND are Infrastructure, Peace, Water, Rural Development and Culture and Recreation.

Keywords: Data analytics; participatory planning; public administration; public policies; text mining; Colombia.

1. Introducción

De acuerdo con el Artículo 342 de la Constitución Política, los intervinientes, trámites y fechas que debe seguir la formulación del Plan Nacional de Desarrollo (PND) se encuentran establecidos en la Ley 152 de 1994, conocida como la ley orgánica del PND. La primera fase se da luego de la elección del presidente de la república, en el mes de agosto, a partir de esa fecha, y con el apoyo del Consejo Nacional de Política Económica y Social (CONPES), se construye las bases del PND, las cuales no integran el PPI, pero sí los principios básicos de la propuesta para iniciar el debate con la sociedad civil. Este debate se realiza desde el 15 de noviembre, a través del Consejo Nacional de Planeación (CNP), que funciona como foro para la discusión del plan por parte de representantes de las entidades territoriales y de los diferentes sectores que lo conforman. El concepto del CNP debe ser entregado a más tardar el 10 de enero al presidente de la república para que se integren sus recomendaciones, si el CONPES lo considera pertinente. Luego, a través del Ministerio de Hacienda y Crédito Público,

el gobierno presentará el PND a consideración del Congreso antes del 7 de febrero, para lo cual convocará a sesiones extraordinarias. El proyecto dará un primer debate en las comisiones económicas de ambas cámaras en sesión conjunta, en máximo 45 días. Así pues, si el trámite se iniciara el 8 de febrero, el primer debate debe surtirse antes del 24 de marzo. Con base en el informe del primer debate, las plenarias de cada una de las cámaras (Senado y Cámara de Representantes) analizarán y decidirán sobre el proyecto presentado en un término no superior a 45 días. Así las cosas, se espera que el PND sea aprobado a más tardar el 7 de mayo, pero si el Congreso no aprueba el PPI en este tiempo, el gobierno podrá ponerlo en vigencia a través de un decreto con fuerza de ley. El PND es el documento que sirve de base y provee los lineamientos estratégicos de las políticas públicas formuladas por el presidente de la república de Colombia, a través de su equipo de gobierno. Su elaboración, socialización, evaluación y seguimiento son responsabilidad directa del Departamento Nacional de Planeación (DNP), el cual se constituye como la unidad rectora de planificación en el país, razón por la cual debe dirigir el proceso de formulación del PND, y coordinar la inclusión de los comentarios pertinentes por parte del CNP, el CONPES y demás actores de la sociedad civil.

El PND 2022-2026 del presidente de la república, Gustavo Petro, se denomina: “Colombia, potencia mundial de la vida”, y representa la hoja de ruta de su gobierno. Este PND tiene como apuesta del gobierno la paz total, se estructura en cinco transformaciones, los actores relevantes para el cambio y, por último, el plan plurianual de inversiones. Las cinco transformaciones son ordenamiento del territorio alrededor del agua y justicia ambiental, seguridad humana y justicia social, derecho humano a la alimentación, internacionalización, transformación productiva para la vida y acción climática y convergencia regional. En la parte de los actores relevantes se identifican los siguientes apartados: el cambio es con las mujeres, Colombia igualitaria, diversa y libre de discriminación, reparación efectiva e integral de las víctimas, nace una nueva generación para la vida y la paz: niñas, niños y adolescentes amados, protegidos e impulsados en sus proyectos de vida con propósito, pueblos y comunidades étnicas, jóvenes con derechos que lideran las transformaciones para la vida y garantías hacia un mundo sin barreras para las personas con discapacidad.

La novedad en este PND fue la estructura. Generalmente, los planes se construían por sectores, lo que hacía más fácil la revisión y el concepto por parte del CNP. Sin embargo, esta nueva estructura generó dos dinámicas interesantes. La primera fue la realización de los diálogos regionales vinculantes por parte del DNP, lo que incitó a la sociedad civil a participar de manera dinámica en los espacios de discusión alrededor del plan. La segunda fue la presentación disruptiva del documento de bases del PND, con cinco transformaciones y varios catalizadores que trataban de presentar una nueva forma de ver el país y proponer los proyectos que serían la hoja de ruta de los próximos cuatro años en términos de política pública en Colombia. En ese contexto, el CNP realizó el IV Encuentro Nacional de Planeación Participativa, con el objetivo de recoger los conceptos de los consejeros territoriales de planeación del país. A este evento asistieron más de 400 consejeros y se recogieron innumerables recomendaciones para el PND. Adicional a esto, se recibieron numerosas propuestas, recomendaciones de diversos sectores, agremiaciones, universidades, organizaciones de la sociedad civil que buscan de alguna manera incluir sus recomendaciones al PND.

Esta cantidad de información sobre el PND y las diversas propuestas y recomendaciones resultaron un desafío para el CNP, que en tiempo récord debe redactar un concepto al plan en pleno periodo decembrino y de vacaciones. Este desafío se relaciona con el contexto contemporáneo en el que se bombardea a los ciudadanos, desde múltiples frentes, con información cuya veracidad y relevancia no son fáciles de identificar. Para los profesionales es difícil mantenerse actualizados en sus campos, y en otros que también son de su interés, debido a las fuentes de información y los generadores de contenido sin fin. Adquirir nueva información útil y veraz resulta una odisea en un ambiente donde, paradójicamente, la información está en la palma de la mano. Distintas industrias han incursionado en el área de la verificación de la información. En este sentido, las redes sociales y los medios de comunicación están a la vanguardia de la verificación de la información de interés general disponible en internet. Sin embargo, documentos de política pública como leyes, decretos y planes gubernamentales no son particularmente atractivos para estos verificadores por sus extensiones. Algunos estudios sobre los PND se han publicado en los últimos años, pero es

reciente el uso de la minería de texto (Arrieta Díaz y Vargas Prieto, 2021; Vargas Prieto y Arrieta Díaz, 2024; Vargas Prieto y Camacho Quintero, 2023; Zapata-Cortés, 2020). Es por esta razón que vale la pena utilizar la minería de texto, aplicada a documentos de política pública, para informar a la ciudadanía sobre los planes que pretenden desarrollar sus gobernantes. Una ciudadanía informada y educada es la mejor arma que tiene la democracia para su defensa.

Este artículo analiza las bases del PND 2022-2026, las preguntas que guiaron la investigación son las siguientes: ¿Cuáles son sus temas más importantes, y cómo se relacionan entre sí? ¿Cuáles son los términos más usados?

El artículo está estructurado en tres partes. En la primera, se presenta la metodología utilizada. En la segunda, se presentan los resultados. En la tercera, se presentan las conclusiones.

2. Metodología

A través de la minería de datos se identifican patrones relevantes dentro de un conjunto de datos de gran tamaño (*big data*) (IBM, 2023). Este proceso de generación de nuevo conocimiento requiere conjuntos de datos ordenados, donde el orden depende de la estructura de la fuente de dichos datos. Cuando la minería de datos se aplica a los textos, minería de texto (IBM, 2023), el análisis está completamente bajo el control de los científicos que lo llevan a cabo, ya que la estructura de los datos la construyen ellos mismos a partir de los textos sujetos de análisis. Cuando la fuente de los datos fue construida por otras personas, o sistemas, ajenos a los analistas, se hace necesario llevar a cabo un proceso de limpieza de los datos, con el fin de crear un conjunto ordenado de datos que se pueda analizar de manera rigurosa. En general, el proceso de minería de datos se desarrolla a través de: la limpieza de los datos, la integración de datos, la selección de datos, la transformación de datos, el descubrimiento de patrones, la evaluación de patrones y la presentación de resultados (Han

et al., 2012). La minería de datos tiene una gran demanda en la sociedad actual, la cual confía cada vez más tareas a los computadores inteligentes y autónomos.

A través de la minería de texto se busca conocer la estructura del PND, así como sus aspectos más relevantes, desde el punto de vista imparcial (en este caso sin sesgos políticos) que ofrecen los algoritmos no-supervisados de la minería de texto (Janasik *et al.*, 2009). Para este fin se utilizó el documento denominado: Bases del PND “Colombia, potencia mundial de la vida 2022-2026”, que fue entregado el 15 de noviembre de 2022 por parte del DNP al CNP. La minería de texto a través de métodos no-supervisados se vale de un tipo de aprendizaje que no necesita la definición *a priori* de etiquetas por parte de los analistas para agrupar los datos, lo cual es una ventaja en las etapas iniciales de un análisis de texto. En las etapas tempranas de un análisis de texto, se necesitan técnicas ágiles para explorar rápidamente muchos caminos que permitan entender *grosso modo* el contenido del texto estudiado. En particular, las técnicas de agrupación aplicadas al documento del PND permiten reunir grandes conglomerados de texto que son similares entre sí (Al-Anazi *et al.*, 2016).

El primer paso en un análisis de texto es tomar el documento completo y aplicarle filtros que permitan eliminar palabras o caracteres que no aportan para la comprensión del contenido. Estos filtros descartan las referencias bibliográficas y los textos de tablas y gráficas. Adicionalmente, se filtran las *stop words*, que son palabras como artículos y preposiciones, entre otras (Hagiwara, 2021). Una vez los documentos han sido filtrados y vectorizados⁴, los algoritmos de agrupamiento procesan el texto y determinan si existen agrupaciones naturales (grupos) en los datos. Particularmente, el modelo BERTopic, de Grootendorst (2022), extrae representaciones de temas coherentes a través de la implementación de un algoritmo basado en clases de frecuencia de documento de frecuencia inversa (TF-IDF, por su sigla en inglés) (Havrlant y Kreinovich, 2017; Trstenjak *et al.*, 2014). Este algoritmo determina la importancia de un término (una palabra) presente en el documento objeto de análisis, y lo clasifica en una de dos categorías: términos importantes y no-importantes. En este caso, la importancia es en relación con el tema o a un tema principal del documento (Zhang *et al.*, 2011). Con base en el TF-IDF, BERTopic genera grupos de textos con modelos

⁴ Para la definición de vectorización, ver por ejemplo: <https://www.kaggle.com/edchen/text-vectorization>

de lenguaje natural previamente entrenados, concatena estos grupos y, finalmente, genera representaciones de temas. Para optimizar el proceso de concatenamiento se debe reducir la dimensionalidad de los grupos resultantes. Sin embargo, estas concatenaciones se usan principalmente para agrupar textos semánticamente similares y no se usan directamente para generar los temas (Grootendorst, 2022).

Adicionalmente, a través de la vectorización de documentos, se establecen secuencias continuas de palabras en el vector de palabras, conocidas como n -gramas, donde n es el número de palabras consecutivas en la secuencia. Estos n -gramas son útiles en el estudio del lenguaje natural para identificar secuencias de palabras relevantes en un documento. Por ejemplo, un bigrama ($n = 2$) o un trigramas ($n = 3$), puede ser el nombre de un cuerpo colegiado ('concejo-municipal'), una entidad estatal ('policía-nacional') o el nombre de un plan estatal ('plan-nacional-desarrollo'), que sean relevantes en el documento. Esta relevancia se establece a través de la frecuencia con la que un n -grama dado aparece en un documento: entre mayor sea la frecuencia de aparición mayor será la relevancia de la secuencia de palabras.

Las representaciones de temas se modelan como función de los textos que componen cada grupo, donde a cada uno se le asigna un tema. En este sentido, se debe saber qué hace que un tema sea único, en función de su distribución de términos agrupados. Para lograr esa clasificación se crea una medida: M , para representar la importancia de una palabra en un texto, de modo que permite una representación de la importancia de un término para un tema dado. Con este fin, el TF-IDF combina dos estadísticas: frecuencia de términos (f) y frecuencia inversa de documento (d). La frecuencia de término modela un texto. Por otro lado, la frecuencia inversa del documento mide qué tanta información provee un término en un texto. En este caso, d es el logaritmo de la razón entre el número de párrafos en el texto y el número de párrafos que contienen un término dado. Estas dos estadísticas se multiplican entre sí: $M = f \cdot d$, y de esa manera se obtiene la medida para representar la importancia de los términos.

El algoritmo M se implementa de manera iterativa, empezando con los grupos primarios de textos, que en el proceso de concatenación están al nivel de oraciones. Por lo tanto, BERTopic, predice las relaciones entre oraciones analizándolas como un todo. Cada iteración en la agrupación de textos produce una medida de importancia para sus términos, de modo que al fusionar las representaciones del tema menos común con el más similar, se logra reducir la cantidad de temas (Grootendorst, 2022). Esta cantidad de temas se puede parametrizar, de acuerdo con los criterios de los analistas. Al final, el modelo entrega una colección de temas que definen la estructura del documento del PND, con base en su composición del lenguaje natural. Estos temas, y los términos que los conforman, permiten conocer el contenido del documento y la importancia que los diferentes términos tienen en su contexto global.

La conexión de los diferentes temas entre sí se explora a través de sus correlaciones y su estructura jerárquica. En general, los algoritmos utilizados para agrupar temas de un texto se pueden dividir en algoritmos aglomerativos basados en particiones y aquellos basados en probabilidades. Los algoritmos aglomerativos fusionan iterativamente temas del documento en grupos, con base en la correlación por pares. Los grupos resultantes se organizan en una jerarquía de grupos. En los algoritmos basados en particiones, el documento se divide en grupos mutuamente excluyentes (no jerárquicos). El proceso de división optimiza la correlación entre los temas del documento dentro de un grupo (Al-Anazi *et al.*, 2016). Los algoritmos aglomerativos clasifican temas correlacionados en función del espacio de características más cercano en el conjunto de entrenamiento; este se determina calculando el coseno del ángulo entre los dos vectores de características creados a partir de la vectorización. Estos vectores están compuestos por las palabras de cada tema del texto analizado, y esto determina su dimensionalidad. Si el coseno del ángulo es cercano a 1, los dos temas están muy correlacionados. Por el contrario, si el coseno del ángulo es cercano a 0, los dos temas no están correlacionados. Los algoritmos aglomerativos combinan sucesivamente los pares de temas más cercanos, hasta que todos los temas forman un grupo grande que define todo el documento. A esta estructura de temas se le denomina jerarquía de grupos (Kavitha y Punithavalli, 2010).

La correlación entre temas del PND se presenta a través de una matriz de correlación, donde cada tema aparece como parte de las filas y de las columnas. La correlación se representa a través de una escala de colores: correlación, donde el azul oscuro implica que dos temas están muy correlacionados; verde claro, que implica que dos temas tienen una correlación baja. La mayor correlación se da entre un tema consigo mismo, correlación igual a 1, sobre la diagonal de la matriz. Un tema que presenta múltiples correlaciones con otros se considera un tema transversal dentro del documento.

La jerarquía de grupos (jerarquía de temas) se presenta a través de un esquema tipo árbol, donde los pares primarios de temas están definidos por una partición jerárquica del espacio multidimensional formado por estos. En el árbol, cada nodo primario es asociado con un rango del dominio multidimensional y la raíz está asociada con todo el documento. El rango de cada nodo interno está asociado con un par de rangos que representan una partición del espacio multidimensional, o con un rango y un tema, como se muestra en la figura 1. La raíz del árbol es la línea más a la derecha del diagrama. La estructura del árbol se obtiene a través de la implementación de un algoritmo que realiza una partición binaria de todo el documento, y luego divide recursivamente cada par de subbloques con el mismo esquema jerárquico binario que se obtiene usando la suma de cuadrados (SSQ) de cada tema. La SSQ de un tema es una función del conjunto de términos que componen un tema: S , sus relaciones entre sí: Q , y el número de temas: N : $SSQ(S, Q, N)$ (Campello *et al.*, 2013). Cada partición binaria busca maximizar la diferencia entre las SSQ de los temas: ΔSSQ . La división recursiva continúa siempre y cuando la ΔSSQ sea mayor que la SSQ promedio del documento. Después de las primeras particiones que producen grandes reducciones de SSQ, los valores de ΔSSQ se vuelven cada vez más pequeños, hasta que después de k particiones, tanto las SSQ como las ΔSSQ se vuelven 0, dado que cada término se ha convertido en su propio tema. Por lo tanto, la reducción promedio de SSQ por partición es: SSQ_0/k , y se debe comparar este valor con la ΔSSQ en cada partición para decidir cuándo se debe parar de dividir.

Figura 1. Definiciones gráficas de los nodos



Nota. Nodo primario, enmarcado en un círculo magenta; nodo interno, enmarcado en un círculo naranja; rango, un par de líneas horizontales azules. Cada línea horizontal, azul primario (a la izquierda), representa un tema.

Fuente. Elaboración propia.

Los resultados del modelo BERTopic aplicado al documento del PND se visualizan utilizando Microsoft Power BI. Esta herramienta de visualización presenta los términos más importantes en el documento, con base en el algoritmo M , en una nube de palabras. Esta visualización permite comparar los resultados de M entre sí a través del tamaño de las palabras en la nube: mayor tamaño implica mayor importancia dentro del documento. Por otro lado, Power BI permite visualizar párrafos ejemplo de cada tema establecido, y muestra todos los párrafos que pertenecen a cada tema. Esto les permite a los analistas tener acceso a los párrafos que se han clasificado por temas. Adicionalmente, Power BI muestra la afinidad entre temas a través de una matriz de correlación (Clarke y Cooke, 1978). Finalmente, la herramienta de visualización permite apreciar la agrupación jerárquica de los temas que definen la estructura del documento con base en el lenguaje natural (Murtagh y Contreras, 2012). De esta estructura jerárquica se puede comprender cómo se piensa ejecutar el PND en su parte conceptual.

Tabla 1. Importancia relativa de los primero 10 términos en la figura 2, se incluye el tema al que pertenece cada término

Término	Importancia relativa	Tema
Colombia	1,00	Colombia
Paz	0,70	Paz
Discapacidad	0,69	Población discapacitada
Género	0,69	Mujeres e inclusión
Alimentos	0,54	Alimentación
Mujeres	0,53	Mujeres
Educación	0,53	Educación y salud
Transporte	0,51	Infraestructura
Nacional	0,50	Colombia, cambio climático e implementación acuerdo de paz
Reparación	0,45	Conflicto armado
Integral	0,43	Educación y salud, conflicto armado y desarrollo rural
Fluviales	0,15	Infraestructura

Fuente. Elaboración propia.

El análisis de minería de texto identificó 23 temas dentro del PND. Estos temas están definidos por medio de términos. El término más importante es Colombia, seguido de paz, discapacidad, género y alimentos. Los dos primeros términos representan los temas con el mismo nombre. Siendo coherente con el eje transversal del PND, la consecución de la paz total en Colombia. Los otros tres términos están dentro de los temas de mujeres e inclusión, alimentación y mujeres nuevamente.

3.2. Bigramas y trigramas con mayor frecuencia en el PND

La relevancia de los bigramas y trigramas se establece con base en una frecuencia relativa para cuantificar el número de apariciones que tuvo cada n -grama en el documento. En ese sentido, la mayor frecuencia relativa es 1, la de los n -gramas más frecuentes, y las demás secuencias de palabras se comparan con estos a través de su frecuencia relativa. Los 20

bigramas (columnas de la izquierda) y los 20 trigramas (columnas de la derecha) más relevantes del PND se presentan en la tabla 2.

Tabla 2. Escalafón con los 20 bigramas y trigramas con mayor frecuencia relativa en el PND

Frecuencia relativa	Bigramas	Frecuencia relativa	Trigramas
1,00	Gobierno-nacional	1,00	Plan-nacional-desarrollo
1,00	Sistema-nacional	0,99	Niños-niñas-adolescentes ⁵
0,76 ⁶	Plan-nacional	0,7	Pueblo-comunidades-étnicas
0,68	Política-pública	0,65	Reforma-rural-integral
0,61	Entidades-territoriales	0,57	Colombia-potencia-mundial
0,58	Derechos-humanos	0,57	Potencia-mundial-vida
0,56	Cambio-climático	0,57	Gestión-riesgo-desastres
0,52	Ordenamiento-territorial	0,49	América-Latina-Caribe
0,5	Enfoque-género	0,49	Derecho-humano-alimentación
0,44	Paz-total	0,49	Sistema-nacional-cuidado
0,42	Nacional-desarrollo	0,49	Inspección-vigilancia-control
0,42	Seguridad-humana	0,49	Ciencia-tecnología-innovación
0,42	Enfoque-diferencial	0,39	Necesidades-cambio-identificadas
0,4	Conflicto-armado	0,39	Trabajo-digno-decente
0,4	Garantía-derechos	0,39	Sexuales-identidades-género
0,35	Toma-decisiones	0,35	Seguridad-humana-justicia
0,34	Persona-discapacidad	0,35	Sistemas-transporte-público
0,34	Economía-popular	0,35	Recreación-actividad-física
0,32	Riesgo-desastres	0,35	Defensoras-derechos-humanos
0,31	Construcción-paz	0,35	Orientaciones-sexuales-identidades

Fuente. Elaboración propia.

⁵ Este trigramas combina la información de dos trigramas que se consideran equivalentes: niños-niñas-adolescentes y niñas-niños-adolescentes.

⁶ En esta posición se excluyen los bigramas: comunidades-étnicas y pueblo-comunidades, ya que la relevancia de esta secuencia de palabras se establece con el trigramas: pueblo-comunidades-étnicas.

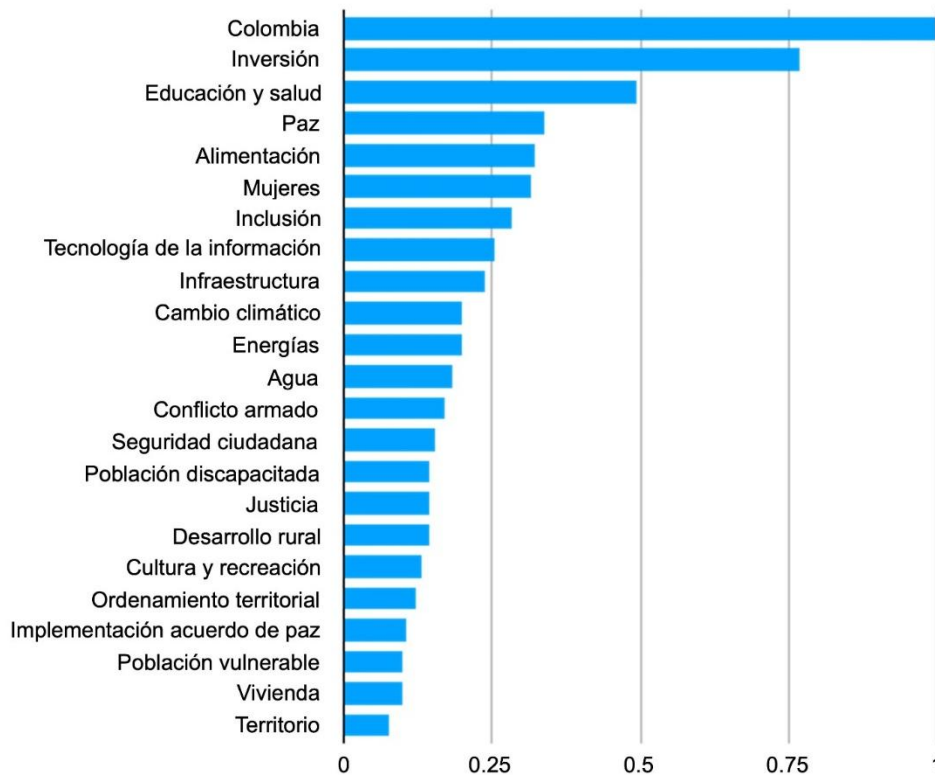
De acuerdo con la tabla 2, el bigrama más frecuente es: gobierno-nacional, mientras que el trigramas más frecuente es: plan-nacional-desarrollo. Esto supone la importancia de la institucionalidad dentro del documento. Los bigramas que continúan en la lista son sistema nacional, plan nacional, política pública, entidades territoriales, derechos humanos, cambio climático, ordenamiento territorial y enfoque de género. Aquí se podría identificar las prioridades del gobierno nacional con respecto al cambio climático y al enfoque territorial del plan con el tema de ordenamiento territorial. También se encuentra el enfoque de género que aparece dentro de los términos más importantes del plan. Con respecto a los trigramas, se encuentran en la lista: pueblo-comunidades-étnicas, reforma-rural-integral, Colombia-potencia-mundial, gestión-riesgo-desastre, América-Latina-Caribe, derechos-humanos-alimentación. En esta parte, se podría identificar como prioridad del gobierno las comunidades étnicas dentro de los actores relevantes del plan y la reforma rural integral como apuesta política, que aparece incluso como bandera del logro de la paz total dentro de las transformaciones.

Si se analiza toda la tabla de bigramas y trigramas, su contenido muestra que, en el PND, la paz, las minorías, la alimentación, la educación, el transporte y los derechos humanos son protagonistas, además de lo mencionado en el párrafo anterior.

3.3. Temas más relevantes del PND

Los 23 temas identificados al usar el modelo BERTopic en el PND se presentan en la siguiente figura:

Figura 3. Frecuencia relativa entre los temas del PND

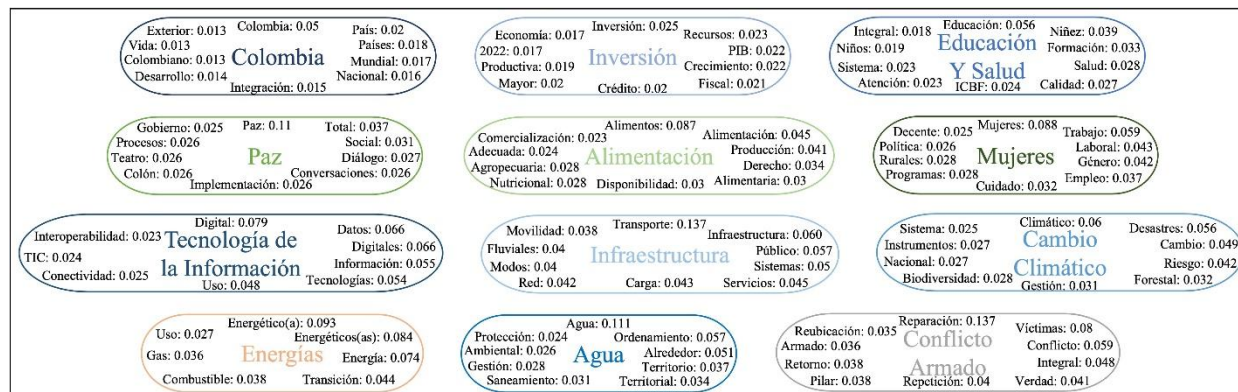


Fuente. Elaboración propia.

Los temas se ordenan por el número de párrafos que los conforman, siendo el tema Colombia el que más agrupa. Con base en este ordenamiento, se define una frecuencia relativa para cuantificar el número de párrafos que le fueron asignados a cada tema. En ese sentido, la mayor frecuencia relativa es 1, la del tema Colombia. Las demás frecuencias relativas son menores, y permiten hacer una comparación cuantitativa del número de párrafos en cada tema, en comparación con el número de párrafos asignados al tema Colombia. El contenido de cada uno de estos temas se puede explorar a partir de los diez términos más importantes que lo articulan.

La figura 4 y la figura 5⁷ presentan cada uno de los 23 temas con sus 10 términos más importantes⁸. El término más importante es el que está encima del nombre de cada tema, y la importancia de los términos decrece en el sentido de las manecillas del reloj.

Figura 3. 12 de los 23 temas identificados en el PND con sus 10 términos más importantes



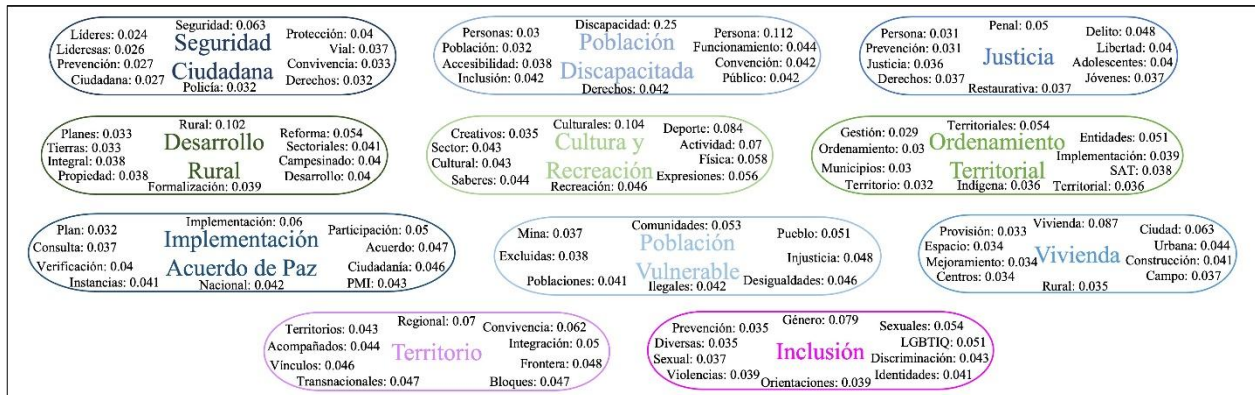
Fuente. Elaboración propia.

El tema Colombia, tiene que ver con el país como territorio, con sus nacionales y con sus relaciones con otros países. El tema inversión, está relacionado con la ejecución de recursos y su impacto en la sociedad. El tema educación y salud, tiene que ver con los planes para mejorar los sistemas educativos y de salud. El tema paz, está relacionado con los procesos de paz que se ejecutarán en el marco del PND, y con los procesos de paz firmados en Colombia en la última década. El tema alimentación, tiene que ver con la producción y distribución de alimentos en Colombia, y cómo estas afectan a la población. El tema mujeres, está relacionado con la búsqueda de la igualdad social con la población masculina. El tema tecnología de la información, se define por sí mismo. El tema infraestructura, tiene que ver con mejorar las condiciones a nivel nacional del transporte, tanto de pasajero como de carga. El tema cambio climático, se define por sí mismo. El tema energías, está relacionado con los planes de transición energética en el país. El tema agua, tiene que ver con el ordenamiento territorial alrededor del agua. El tema conflicto armado, está relacionado con las consecuencias sobre la población civil que tienen los diversos conflictos con grupos ilegales que afronta el estado.

⁷ El término SAT en el tema ordenamiento territorial se refiere al sistema de administración del territorio. El término PMI en el tema implementación acuerdo de paz se refiere al plan marco de implementación.

⁸ En unos pocos temas no existen 10 términos con suficiente importancia para ser mostrados.

Figura 4. 11 de los 23 temas identificados en el PND con sus 10 términos más importantes



Fuente. Elaboración propia.

El tema seguridad ciudadana, se define por sí mismo. El tema población discapacitada, tiene que ver con atender las necesidades de la población en condición de discapacidad. El tema justicia, está relacionado con los delitos que ocurren alrededor de la ciudadanía, particularmente la joven. El tema desarrollo rural, se define por sí mismo. El tema cultura y recreación, se define por sí mismo. El tema ordenamiento territorial, se define por sí mismo. El tema implementación acuerdo de paz, tiene que ver con la implementación del acuerdo de paz firmado por el estado y por las FARC. El tema población vulnerable, está relacionado con la situación de las comunidades étnicas en el país. El tema vivienda, tiene que ver con brindar viviendas dignas a la población que las necesita. El tema: territorio, está relacionado con aspectos propios de las regiones fronterizas del país. Finalmente, el tema inclusión, tiene que ver con promover la igualdad de género en el país. Estas definiciones pretenden brindar una orientación rápida sobre los aspectos tratados con más relevancia en el PND. Sin embargo, los detalles de cada tema, y su relación con otros temas, se presentan a continuación.

Tabla 3. Detalles de cada tema y su relación con otros temas

Temas por Z	Z	Temas por M del PT	M	Temas por E	E
Población discapacitada	0,670	Población discapacitada	0,250	Población discapacitada	1
Cultura y recreación	0,583	Infraestructura	0,137	Infraestructura	2
Conflicto armado	0,552	Conflicto armado	0,137	Conflicto armado	2
Infraestructura	0,551	Agua	0,111	Cultura y recreación	3
Territorio	0,503	Paz	0,110	Desarrollo rural	4
Tecnología de la información	0,462	Cultura y recreación	0,104	Agua	5
Desarrollo rural	0,459	Desarrollo rural	0,102	Tecnología de la información	6
Inclusión	0,455	Mujeres	0,088	Territorio	7
Vivienda	0,442	Alimentación	0,087	Vivienda	7
Implementación acuerdo de paz	0,437	Vivienda	0,087	Inclusión	7
Población vulnerable	0,428	Inclusión	0,079	Mujeres	8
Agua	0,423	Tecnología de la información	0,079	Paz	9
Energías	0,421	Energías	0,074	Energías	10
Mujeres	0,407	Territorio	0,070	Implementación acuerdo de paz	11
Justicia	0,386	Seguridad ciudadana	0,063	Alimentación	11
Cambio climático	0,377	Cambio climático	0,060	Población vulnerable	12
Ordenamiento territorial	0,374	Implementación acuerdo de paz	0,060	Cambio climático	13
Alimentación	0,369	Educación y salud	0,056	Seguridad ciudadana	14
Paz	0,361	Ordenamiento territorial	0,054	Ordenamiento territorial	15
Seguridad ciudadana	0,339	Población vulnerable	0,053	Justicia	16
Educación y salud	0,290	Colombia	0,050	Educación y salud	17
Inversión	0,206	Justicia	0,050	Colombia	18
Colombia	0,191	Inversión	0,025	Inversión	19

Nota. Centro: temas ordenados de mayor a menor *M* en su primer término (PT). Derecha: temas ordenados de mayor a menor especificidad (E). Izquierda: temas ordenados de mayor a menor *Z*.

Fuente. Elaboración propia.

La importancia de los términos que definen cada tema se debe tener en cuenta para comprender su estructura individual. En este sentido, es interesante mirar el valor de la importancia del término que aparece justo arriba de cada tema (figura 4 y figura 5). Este

término es el más importante para cada tema. Así mismo, el valor de la importancia de ese primer término en cada tema se debe comparar con sus equivalentes en los otros temas, para determinar cuáles son más específicos en su estructura. Un valor menor a 0,08 en la importancia del primer término es un primer indicador de que el tema no es muy específico. En este sentido, un tema poco específico puede abarcar muchos aspectos, o agrupar diversos aspectos alrededor de un eje central. Por otro lado, la suma de las importancias Z , de los 10 términos de cada tema ayuda a clasificar qué tan específico es un tema, en comparación con los demás.

El menos importante de los primeros términos de cada tema es inversión, así: inversión: $M = 0.025$. Adicionalmente, $Z = 0.206$, la segunda menor Z de todas. Lo poco específico de este tema se refleja en el hecho de que abarca muchos aspectos de la inversión que planea realizar el gobierno; la cual tiene propósitos diversos. En los párrafos que tratan el tema inversión se habla de: exportaciones de hidrocarburos y minería, la participación de la industria en el PIB, los PIB de otros países, el impacto del gasto público en la sociedad, la financiación de la inversión que se propone en el PND y su ajuste a la regla fiscal y los sectores que serán objeto de la inversión. Lo poco específico de este tema no es un indicador de poca relevancia de este en el PND. Al contrario, el tema es el segundo con más frecuencia relativa en el documento. La variedad de aspectos tratados y el número de párrafos que se cobijan son indicadores de que el tema es de los más relevantes en el PND. En relación con lo anterior, el tema Colombia tiene la menor Z de todos, siendo el menos específico, pero también es de los temas más relevantes. Este tema establece el escenario en el cual se desarrolla el PND.

Por otro lado, el tema más específico de todos es: población discapacitada. Su primer término, discapacidad, es el más importante de todos los primeros, y su Z es la mayor de todas. En el contexto del tema se promete formular estrategias para empoderar a las personas con discapacidad en aspectos de participación ciudadana, electoral y política; se indicará en el Plan Nacional de Accesibilidad para garantizar el acceso al entorno físico, al transporte, a la información y a las comunicaciones, y se tratarán aspectos relacionados con el ejercicio de los derechos de las personas con discapacidades. En ese sentido, el tema es específico, ya que gira alrededor de beneficiar a un grupo concreto de la población; y tiene un solo propósito

central. Este tema ocupa el puesto 15 en la frecuencia relativa de la figura 3, pero su término discapacidad tiene la tercera importancia relativa del PND, según la tabla 1. Por lo tanto, lo específico del tema no le resta relevancia dentro del PND.

Como se puede ver en la tabla 3, los temas más específicos giran alrededor de uno o pocos propósitos centrales. Esto implica que se utilizan relativamente pocos términos importantes para articular las ideas. Al examinar la figura 4 y figura 5, se evidencia que los temas más específicos tienen importancias en sus términos mayores, más que la importancia promedio. En ese sentido, la importancia de los términos, diferentes al primero de cada tema, tiene un promedio de 0,037. Los temas menos específicos mencionan muchos términos para articular sus ideas, lo cual disminuye la importancia de sus términos más importantes. Para los temas menos específicos se observa en la figura 4 y figura 5 que la mayoría de sus términos tienen una importancia menor al promedio. En el caso del tema justicia, con una especificidad intermedia, tres de sus términos tienen una importancia menor al promedio, tres de sus términos tienen una importancia promedio y tres de sus términos tienen una importancia superior al promedio⁹, lo que justifica su puesto en la escala de especificidad: 16.

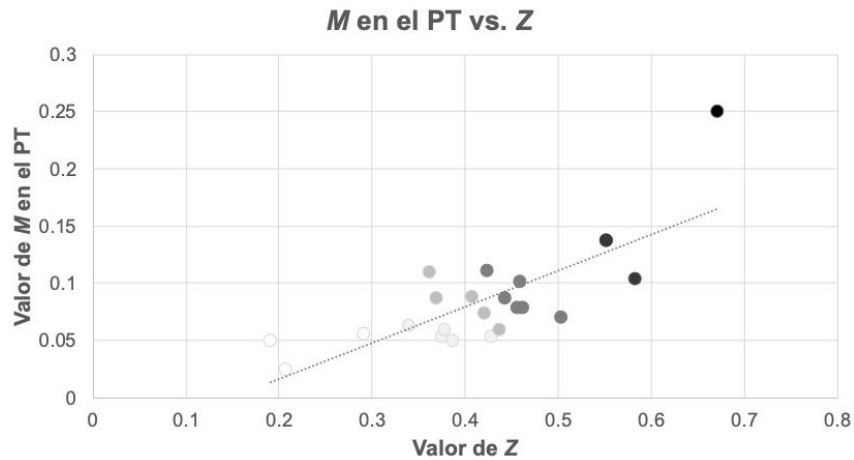
3.4. Relevancia de los temas: combinación de especificidad y frecuencia relativa

Con base en la información combinada de Z y la M del primer término para cada tema, se puede establecer qué tan específico es un tema en particular. La tabla 3 muestra los temas ordenados: de mayor a menor Z (columnas de la izquierda), y de mayor a menor M del primer término (columnas del centro). Lo específico de un tema se puede entonces establecer mirando la correlación que existe entre las Z y las M del primer término. La figura 6 muestra que existe una correlación entre los dos valores para cada tema, y esa correlación se usa para establecer qué tan específico es cada tema. Los puntos más oscuros corresponden a los temas más específicos, y los puntos más claros corresponden a los menos específicos. Ese orden de especificidad se presenta en la tabla 3, en las columnas de la derecha. Se presenta el caso de que algunos temas comparten la misma especificidad. Al revisar la relación de la

⁹ Este argumento excluye la importancia del primer término de cada tema.

especificidad con la frecuencia relativa de cada uno de los temas, no se encuentra una correlación entre las dos variables.

Figura 5. Correlación entre las Z y las M del primer término (PT) de cada tema

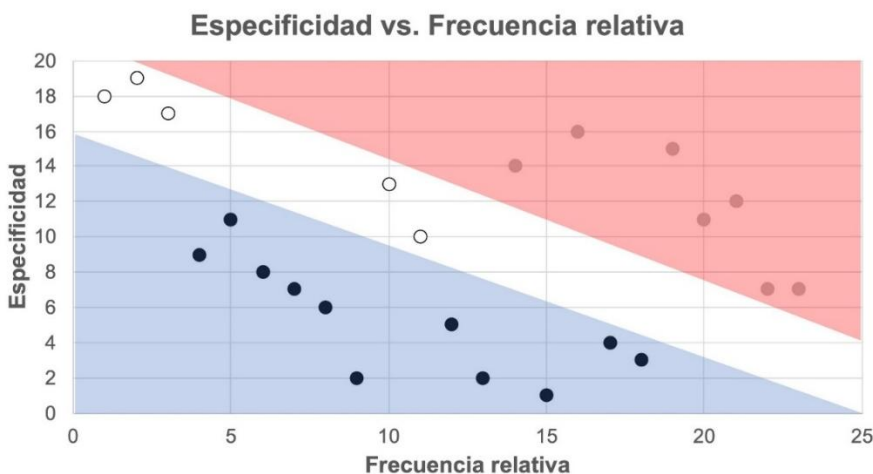


Nota. Los puntos más oscuros son los temas más específicos y los más claros los menos específicos.

Fuente. Elaboración propia.

Sin embargo, la relevancia de los temas identificados dentro del PND se puede cuantificar por zonas:

Figura 6. La combinación de especificidad y frecuencia relativa muestra dos zonas que permiten clasificar la relevancia de los temas



Fuente. Elaboración propia.

La zona azul contiene los temas más relevantes en el PND y la zona roja contiene los temas con menor relevancia. La transición entre las dos zonas contiene los temas de Colombia, inversión, educación y salud, energía y cambio climático, cuya relevancia es ambigua. Estos temas se clasifican con relevancia media y se ubican en el puesto 9 en la tabla 4.

Tabla 4. Relevancia de los temas en el PND

Relevancia	Tema
1	Infraestructura
2	Paz
3	Mujeres
3	Inclusión
3	Tecnología de la información
4	Conflicto armado
5	Alimentación
5	Población discapacitada
6	Agua
7	Desarrollo rural
8	Cultura y recreación
9	Colombia
9	Educación y salud
9	Inversión
9	Energías
9	Cambio climático
10	Seguridad ciudadana
11	Vivienda
12	Territorio
13	Implementación acuerdo de paz
14	Justicia
15	Población vulnerable
16	Ordenamiento territorial

Nota. Los temas 1 a la 8 son los más relevantes, los temas 9 tienen relevancia media, y los temas con relevancia mayor a 9 son los menos relevantes.

Fuente. Elaboración propia.

Hay más temas considerados más relevantes que aquellos considerados menos relevantes. Los temas más relevantes se ubican del puesto 1 al 8 en la tabla 4. Algunos temas comparten la misma relevancia. En esta idea de relevancia es más conveniente observar que hay dos zonas que dan el sentido de relevancia dentro del PND a cada tema, que establecer una jerarquía de relevancia entre los temas de una misma zona. Para complementar la idea de los temas más relevantes se puede comparar el orden de la tabla 4 con los temas presentados en la tabla 1, y se observa que hay una coincidencia en la mayoría de los temas en las dos instancias.

El tema de educación y salud es poco específico, ya que combina dos grandes ideas. Sus términos más importantes están relacionados con la niñez, los sistemas de educación y la salud, la calidad de los servicios que estos prestan, y con la formación de profesionales. La combinación de estos dos aspectos tan relevantes para una sociedad en un solo tema llama la atención. Sin embargo, el hecho de que ambas ideas giren alrededor de aspectos centrales como: sistemas, calidad, atención, niñez e integral, explica por qué se creó este tema. Adicionalmente, las dos ideas comparten un núcleo central: el cuidado de diferentes grupos de la sociedad civil. Este tema, en su aspecto de salud abarca: la optimización del modelo de salud con enfoque promocional y preventivo basado en la atención primaria en salud (APS), la efectividad de la APS a través de la reorientación de los servicios de salud, el fortalecimiento y el aseguramiento de la salud para el cuidado integral de toda la población, la sostenibilidad de los recursos en salud, el enfoque preventivo y predictivo de la salud y el bienestar para juventudes, especialmente para aquellos adolescentes y jóvenes que no estudian ni trabajan, y el aumento de la cobertura de la ruta integral de atención y mantenimiento en salud, entre otros.

El tema, en su aspecto de educación, abarca: un nuevo plan de infraestructura educativa, la educación de calidad para reducir la desigualdad, la resignificación de la jornada escolar, el fortalecimiento de la formación de docentes, directivos docentes y orientadores escolares, la educación media para la construcción de proyectos de vida, la erradicación del analfabetismo y el cierre de inequidades, la educación superior como un derecho, el sistema nacional de cualificaciones (SNC), la cobertura universal del programa de alimentación escolar (PAE), el

impulso a las áreas STEAM¹⁰, la ampliación de la cobertura educativa y la educación con pertinencia para la población campesina, entre otros.

Los dos aspectos centrales del tema educación y salud se encuentran en el cuidado de la población, abarcando: la búsqueda activa de menores con discapacidad para garantizar su bienestar social, las búsquedas activas de jóvenes desescolarizados para promover su acceso a la oferta institucional, el programa de vigías del patrimonio, la transformación institucional del ICBF, el rol estratégico del ICBF en la puesta en marcha del PND, el uso de la medida de pobreza en la niñez para la priorización de políticas públicas para el desarrollo integral de la niñez y la familia, en el marco del Sistema Nacional de Bienestar Familiar (SNBF), el posicionamiento y la consolidación del SNBF articulados con el sistema de protección social universal y adaptativo (SPS), el sistema de cuidado, el sistema nacional de gestión del riesgo de desastres, el sistema de información integrado para la identificación, registro y caracterización del trabajo infantil y sus peores formas, el fortalecimiento del vínculo de las cajas de compensación familiar con FONIÑEZ, y la universalización de la atención integral con calidad, entre otros. Toda esta variedad de aspectos en el tema de educación y salud hacen que su especificidad sea baja y su relevancia sea media. Sin embargo, es un tema transversal en el PND.

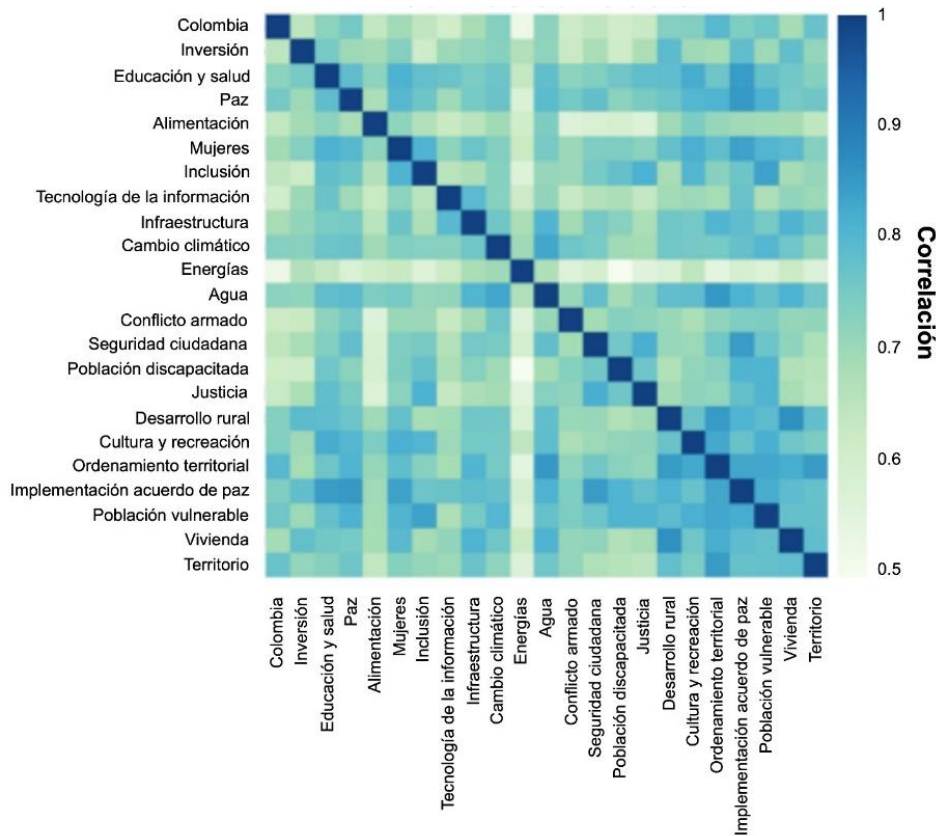
3.5. Correlación de los temas

Las correlaciones entre los 23 temas del PDN se pueden ver en la figura 8. La correlación más fuerte se da entre los temas vivienda y desarrollo rural (correlación: 0,88). Ambos temas coinciden en la participación efectiva y diferencial para la toma de decisiones sobre el desarrollo ambiental, con base en un modelo de ordenamiento que promueva la sostenibilidad ambiental. Así mismo, coinciden en tratar la democratización de la tierra en espacio urbano y rural, y en la formalización, regularización y adjudicación de la propiedad. Coinciden en estrategias para la provisión y mejoramiento de viviendas en los centros poblados y rurales, por ejemplo, permeando los procesos sectoriales y aprovechando las sinergias que resultan de los vínculos urbano-rurales. Otros pares de temas fuertemente

¹⁰ Para conocer la definición de las áreas STEAM ver: <https://erasmuslearn.eu/steam-education-creating-fun-learning-environment-oriented-reality/>

correlacionados entre sí son: agua y ordenamiento territorial, ordenamiento territorial, territorio y paz e implementación del acuerdo de paz (los 3 pares con correlación: 0,87).

Figura 7. Matriz de Correlación del PND



Nota. La escala de colores representa correlación entre dos temas dados. Azul oscuro (cercana a 1) implica gran correlación, verde claro (cercana a 0,5) implica correlación baja.

Fuente. Elaboración propia.

Las correlaciones más fuertes se muestran en la tabla 5, donde también se observan las correlaciones más débiles.

Tabla 5. Pares de temas con correlaciones más fuertes (izquierda) y más débiles (derecha)

Correlaciones más fuertes	Correlaciones más débiles
Vivienda-desarrollo rural	Energía-Colombia
Ordenamiento territorial-agua	Energía-justicia
Territorio-ordenamiento territorial	Alimentación-justicia
Implementación acuerdo de paz-paz	Energía-inclusión
Ordenamiento territorial-desarrollo rural	Energía-desarrollo rural
Implementación acuerdo de paz-educación y salud	Energía-ordenamiento territorial
Implementación acuerdo de paz-seguridad ciudadana	Alimentación-seguridad ciudadana
Ordenamiento territorial-población vulnerable	Energía-paz
Implementación acuerdo de paz-ordenamiento territorial	Alimentación-conflicto armado
Implementación acuerdo de paz-mujeres	Energía-conflicto armado
Población vulnerable-inclusión	Energía-población vulnerable
Ordenamiento territorial-vivienda	Energía-territorio

Fuente. Elaboración propia.

Los temas con el mayor número de correlaciones fuertes con otros temas son: implementación acuerdo de paz y población vulnerable (18 correlaciones por encima de 0,75). Otros temas con múltiples correlaciones fuertes (17 correlaciones por encima de 0,75) con otros temas son: educación y salud y paz. Estos cuatro últimos temas son los más transversales en el PND. Otros temas transversales en el PND son (entre 15 y 13 correlaciones por encima de 0.75): infraestructura, cambio climático, agua, desarrollo rural, cultura y recreación, ordenamiento territorial y vivienda. Por otro lado, el tema menos transversal en el PND es energía, sin correlaciones por encima de 0,75. Su correlación más alta, 0,69, es con el tema cambio climático. Otros temas poco transversales en el PND son: alimentación, conflicto armado y tecnología de la información. La tabla 6 presenta el escalafón de temas por mayor número de correlaciones fuertes con otros temas.

Tabla 6. Escalafón de temas por número de correlaciones fuertes con otros temas

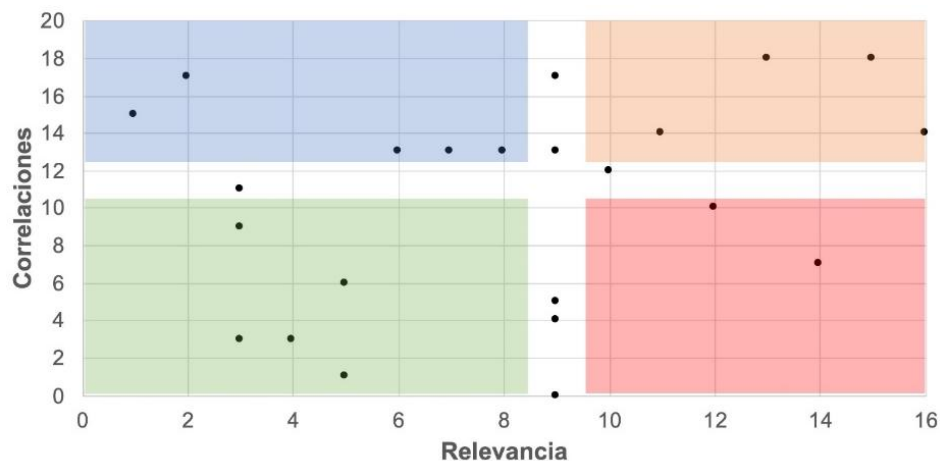
Escalafón de correlaciones	Tema
1	Implementación acuerdo de paz
1	Población vulnerable
2	Educación y salud
2	Paz
3	Infraestructura
4	Ordenamiento territorial
4	Vivienda
5	Cambio climático
5	Agua
5	Desarrollo rural
5	Cultura y recreación
6	Seguridad ciudadana
7	Mujeres
8	Territorio
9	Inclusión
10	Justicia
11	Población discapacitada
12	Inversión
13	Colombia
14	Tecnología de la información
14	Conflicto armado
15	Alimentación
16	Energías

Nota. Se presenta el orden de mayor a menor número de correlaciones fuertes. Hay temas que ocupan el mismo puesto en el escalafón.

Fuente. Elaboración propia.

Al revisar la relación entre lo transversal de un tema en el PND y su relevancia en este, no se encuentra una correlación entre las dos variables. Sin embargo, esta relación se puede cuantificar por zonas:

Figura 8. Correlaciones vs. relevancia



Nota. Escalafón de correlaciones fuertes (mayores a 0,75) en función de la relevancia de cada tema. Temas fundamentales en el PND: zona azul. Temas especiales en el PND: zona verde. Temas particulares en el PND: zona naranja. Temas poco trascendentales en el PND: zona roja.

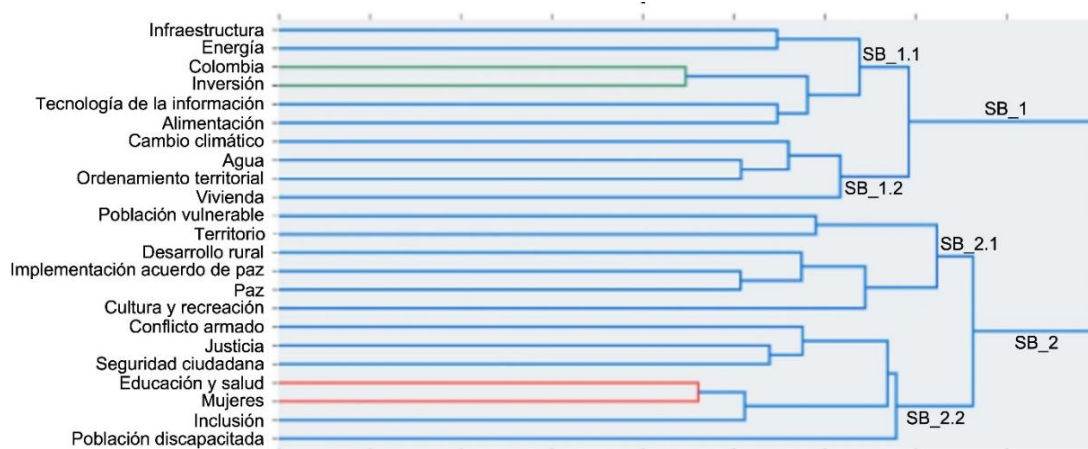
Fuente. Elaboración propia.

Un tema que es transversal en el PND, con alto número de correlaciones fuertes con otros temas, y además es de los más relevantes, se cataloga como fundamental, esta es la zona azul. Los temas fundamentales del PND son: infraestructura, paz, agua, desarrollo rural y cultura y recreación. Por otro lado, un tema que es poco transversal en el PND, con pocas correlaciones fuertes con otros temas, y además es de los más relevantes, se cataloga como especial, esta es la zona verde. Los temas especiales son: inclusión, tecnología de la información, conflicto armado, alimentación y población discapacitada. De estos, cuatro son de los más específicos, y uno tiene una especificidad media. Estos temas son especiales porque son relevantes en el PND, pero no son transversales dada su naturaleza específica. En la zona naranja aparecen temas transversales, pero con relevancia menor, los cuales se catalogan como particulares. Los temas particulares son: vivienda, implementación acuerdo de paz, población vulnerable y ordenamiento territorial. Finalmente, en la zona roja, aparecen dos temas poco trascendentales, que no son transversales y tienen una relevancia menor en el PND: territorio y justicia. Los temas con relevancia media y los temas: mujeres y seguridad ciudadana quedan por fuera de esta clasificación.

3.6. Jerarquía de los temas

La jerarquía de los temas en el PND se presenta en la figura 10. Esta jerarquía optimiza las relaciones entre nodos. Esta optimización tiene como una de sus bases a los nodos primarios de los temas: paz-implementación acuerdo de paz, justicia-seguridad ciudadana, mujeres-educación y salud y población vulnerable-territorio, cuyas correlaciones son mayores que el 77 % de todas las correlaciones entre temas. Estos nodos primarios soportan el segundo subbloque (SB_2), que está compuesto por temas de carácter social, relacionados con la ciudadanía. El subbloque SB_2.1, de este segundo subbloque, está relacionado principalmente con los temas: paz, implementación acuerdo de paz, población vulnerable, desarrollo rural y cultura y recreación. El subbloque SB_2.2, de este segundo subbloque, tiene tres bases principales, una compuesta por los temas: justicia, seguridad ciudadana y conflicto armado, otra compuesta por los temas: mujeres, inclusión y población discapacitada, y la última compuesta por el tema de educación y salud.

Figura 9. Jerarquía de temas en el PND



Nota. Se indican los dos subbloques más grandes en el documento: SB_1 y SB_2, y cada uno de sus dos subbloques más grandes: SB_1.1, SB_1.2, SB_2.1 y SB_2.2.

Fuente. Elaboración propia.

Las bases de los nodos primarios del primer subbloque (SB_1) son los temas: infraestructura-energía, Colombia-inversión, agua-ordenamiento territorial y alimentación - tecnología de la información. Este primer subbloque tiene una composición de temas relacionados con recursos económicos, tecnológicos y naturales. El subbloque SB_1.1, del subbloque SB_1, está relacionado principalmente con los temas de infraestructura, energía y tecnología de la información. Las correlaciones de los nodos primarios de este subbloque son menores que el 94 % de todas las correlaciones entre temas, con un promedio de 0,66. El subbloque SB_1.2, del subbloque SB_1, está relacionado con los temas de cambio climático, agua, ordenamiento territorial y vivienda. En este caso, las correlaciones de los nodos primarios del subbloque SB_1.2 son fuertes, con un promedio de 0,81, con todas las correlaciones mayores que el 63 % de todas las correlaciones entre temas.

4. Conclusiones

La hoja de ruta que cada gobierno nacional de Colombia ejecuta es su Plan Nacional de Desarrollo. Estos planes normalmente son presentados por los líderes del gobierno, y la presentación tiene tintes políticos. Sin embargo, dada la importancia que un PND tiene para una sociedad, vale la pena conocerlo sin sesgo político, ni social. Una forma de lograr esto es extraer la información contenida en un PND a través de la minería de texto. Los principales resultados se presentan a continuación.

El análisis de minería de texto identificó 23 temas dentro del PND. Estos temas están definidos por medio de términos. El término más importante es Colombia, seguido de paz, discapacidad, género y alimentos. Los dos primeros términos representan los temas con el mismo nombre. Siendo coherente con el eje transversal del PND, la consecución de la paz total en Colombia. Los otros tres términos están dentro de los temas de mujeres e inclusión, alimentación y mujeres, nuevamente.

El análisis permitió la identificación de los bigramas y trigramas con mayor frecuencia en el PND. De acuerdo con la tabla 2, el bigrama más frecuente es: gobierno-nacional, mientras que el trigramas más frecuente es: plan-nacional-desarrollo. Esto supone la importancia de la institucionalidad dentro del documento. Los bigramas que continúan en la lista son: sistema

nacional, plan nacional, política pública, entidades territoriales, derechos humanos, cambio climático, ordenamiento territorial y enfoque de género. Aquí se podría identificar las prioridades del gobierno nacional con respecto al cambio climático y al enfoque territorial del plan con el tema ordenamiento territorial. También se encuentra el enfoque de género que aparece dentro de los términos más importantes del plan. Con respecto a los trigramas, se encuentran en la lista: pueblo-comunidades-étnicas, reforma-rural-integral, Colombia-potencia-mundial, gestión-riesgo-desastre, América-Latina-Caribe, derechos-humanos-alimentación. En esta parte, se podría identificar como prioridad del gobierno las comunidades étnicas dentro de los actores relevantes del plan, y la reforma rural integral como apuesta política que aparece incluso como bandera del logro de la paz total dentro de las transformaciones.

Si se analiza toda la tabla de bigramas y trigramas, su contenido muestra que, en el PND, la paz, las minorías, la alimentación, la educación, el transporte y los derechos humanos son protagonistas, además de lo mencionado en el párrafo anterior.

De los 23 temas identificados al usar el modelo BERTopic en el PND se presentan tres que se denotan más que los demás: Colombia, inversión y educación y salud. El tema: Colombia, tiene que ver con el país como territorio, con sus nacionales y con sus relaciones con otros países. El tema: inversión, está relacionado con la ejecución de recursos y su impacto en la sociedad. El tema: educación y salud, tiene que ver con los planes para mejorar los sistemas educativos y de salud. Seguidos de paz, alimentación, mujeres e inclusión. El tema: paz, está relacionado con los procesos de paz que se ejecutarán en el marco del PND y con los procesos de paz firmados en Colombia en la última década. El tema: alimentación, tiene que ver con la producción y distribución de alimentos en Colombia y cómo estas afectan a la población. El tema: mujeres, está relacionado con la búsqueda de la igualdad social con la población masculina. El tema: inclusión, tiene que ver con promover la igualdad de género en el país.

El término menos importante de los primeros de cada tema es inversión, lo poco específico de este tema se refleja en el hecho de que abarca muchos aspectos de la inversión que planea realizar el gobierno; la cual tiene propósitos diversos. En los párrafos que tratan el tema

inversión se habla de: exportaciones de hidrocarburos y minería, la participación de la industria en el PIB, los PIB de otros países, el impacto del gasto público en la sociedad, la financiación de la inversión que se propone en el PND y su ajuste a la regla fiscal y los sectores que serán objeto de la inversión. Lo poco específico de este tema no es un indicador de poca relevancia de este en el PND. Al contrario, el tema es el segundo con más frecuencia relativa en el documento. La variedad de aspectos tratados y el número de párrafos que se cobijan bajo este tema son indicadores de que es de los más relevantes en el PND. En relación con lo anterior, el tema Colombia tiene la menor *Z* de todos, siendo el menos específico, pero también es de los temas más relevantes, ya que establece el escenario en el cual se desarrolla el PND.

Por otro lado, el tema más específico de todos es población discapacitada. Su primer término, discapacidad, es el más importante de todos los primeros términos, y su *Z* es la mayor de todas. En el contexto del tema se promete formular estrategias para empoderar a las personas con discapacidad en aspectos de participación ciudadana, electoral y política; también formular el plan nacional de accesibilidad para garantizar acceso al entorno físico, al transporte, a la información y las comunicaciones, y se tratan aspectos relacionados con el ejercicio de los derechos de las personas con discapacidades. En ese sentido, el tema es específico, ya que gira alrededor de beneficiar a un grupo concreto de la población; tiene un solo propósito central.

Con respecto a la relevancia de los temas, se observa que hay una coincidencia en la mayoría de los temas en los primeros resultados, con la combinación de especificidad y frecuencia relativa, encontrando que el tema de educación y salud es poco específico, ya que combina dos grandes ideas. Sus términos más importantes están relacionados con la niñez, los sistemas de educación y salud, y la calidad de los servicios que estos prestan, y con la formación de profesionales. La combinación de estos dos aspectos tan relevantes para una sociedad en un solo tema llama la atención, sin embargo, el hecho de que ambas ideas giren alrededor de aspectos centrales como: sistemas, calidad, atención, niñez e integral, explica por qué está creado el tema.

Los dos aspectos centrales del tema educación y salud se encuentran en una variedad de aspectos que hacen que su especificidad sea baja y su relevancia sea media. Sin embargo, es un tema transversal en el PND.

Al revisar la correlación de los temas se observa que los temas fundamentales del PND son: infraestructura, paz, agua, desarrollo rural y cultura y recreación. Los temas especiales son: inclusión, tecnología de la información, conflicto armado, alimentación y población discapacitada. Estos temas son especiales, porque son relevantes en el PND, pero no son transversales dada su naturaleza específica. Los temas particulares son: vivienda, implementación acuerdo de paz, población vulnerable y ordenamiento territorial. Hay dos temas poco trascendentales, que no son transversales y tienen una relevancia menor en el PND son: territorio y justicia.

Con respecto a las limitaciones de la metodología, se puede señalar que la minería de texto se realiza sobre un documento en un momento único y no dinámico, por lo que en el caso del análisis de un PND no se puede evaluar qué hace el gobierno con el contenido del texto analizado.

Finalmente, en el ámbito de la administración y la gestión pública, entender las propuestas y prioridades de un PND es fundamental para diseñar e implementar políticas efectivas que impulsen el desarrollo económico, social y ambiental. El artículo contribuye a este campo al ofrecer una metodología y herramientas específicas, como la minería de texto, para analizar de manera sistemática y estructurada la información contenida en documentos complejos como los PND. El aporte real de nuevo conocimiento radica en la aplicación exitosa de la analítica de datos al ámbito de la formulación de políticas públicas, identificando no solo los temas más relevantes, sino también su interrelación y su alineación con los objetivos estratégicos del plan. Esto proporciona a los tomadores de decisiones, investigadores y profesionales en el área de la administración y la gestión pública información valiosa para la toma de decisiones informadas y la elaboración de estrategias efectivas de desarrollo nacional.

5. Referencias

- Al-Anazi, S., AlMahmoud, H. & Al-Turaiki, I. (2016). Finding similar documents using different clustering techniques. *Procedia Computer Science*, 82, 28-34.
<https://doi.org/10.1016/j.procs.2016.04.005>
- Arrieta Díaz, E. y Vargas Prieto, A. (2021). Analítica de datos aplicada al Plan Nacional de Desarrollo de 2018-2022, "Pacto por Colombia, pacto por la equidad". *Estudios Políticos*, (62), 215-239.
<https://doi.org/10.17533/udea.espo.n62a09>
- Campello, R. J., Moulavi, D. & Sander, J. (2013, April 14-17). Density-based clustering based on hierarchical density estimates. In J. Pei, V. S. Tseng, L. Cao, H. Motoda, G. Xu (Eds.), *Advances in Knowledge Discovery and Data Mining* (pp. 160-172). 17th Pacific-Asia Conference, PAKDD, Gold Coast, Australia, Proceedings, Part II. https://doi.org/10.1007/978-3-642-37456-2_14
- Clarke, G. M. & Cooke, D. (1978). *A basic course in statistics* (second edition).
- Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. Arxiv.
<https://arxiv.org/abs/2203.05794>
- Hagiwara, M. (2021). *Real-world natural language processing: practical applications with deep learning*. Manning Publications.
- Han, J., Kamber, M. & Pei, J. (2012). *Data mining: Concepts and techniques* (third edition).
<https://doi.org/10.1016/C2009-0-61819-5>
- Havrlant, L. & Kreinovich, V. (2017). A simple probabilistic explanation of term frequency-inverse document frequency (TF-IDF) heuristic (and variations motivated by this explanation). *International Journal of General Systems*, 46(1), 27-36.
<https://doi.org/10.1080/03081079.2017.1291635>
- IBM. (2023). *What is data mining?* <https://www.ibm.com/topics/data-mining>
- Janasik, N., Honkela, T. & Bruun, H. (2009). Text mining in qualitative research: Application of an unsupervised learning method. *Organizational Research Methods*, 12(3), 436-460.
<https://doi.org/10.1177/1094428108317202>

- Kavitha, V. & Punithavalli, M. (2010). Clustering time series data stream - a literature survey. *International Journal of Computer Science and Information Security*, 8(1), 289-294.
- Murtagh, F. & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(1), 86-97.
<https://doi.org/10.1002/widm.53>
- Trstenjak, B., Mikac, S. & Donko, D. (2014). KNN with TF-IDF based framework for text categorization. *Procedia Engineering*, 69, 1356-1364. <https://doi.org/10.1016/j.proeng.2014.03.129>
- Vargas Prieto, A. y Arrieta Díaz, E. (2024). Identificación y caracterización de los consejos territoriales de planeación en Colombia. *Tendencias*, 25(1), 109-138.
<https://doi.org/10.22267/rtend.242501.243>
- Vargas Prieto, A. y Camacho Quintero, M. A. (2023). *El Plan Nacional de Desarrollo 2022-2026 desde el Consejo Nacional de Planeación*. Departamento Nacional de Planeación.
<https://www.dnp.gov.co/publicaciones/Planeacion/Paginas/el-plan-nacional-de-desarrollo-2022-2026-desde-el-consejo-nacional-de-planeacion.aspx>
- Zapata-Cortés, O. L. (2020). Reflexión sobre los planes de desarrollo en Colombia. *Bitácora Urbano Territorial*, 30(3), 233-246. <https://doi.org/10.15446/bitacora.v30n3.86811>
- Zhang, W., Yoshida, T. & Tang, X. (2011). A comparative study of TF*IDF, LSI and multi-words for text classification. *Expert Systems with Applications*, 38(3), 2758-2765.
<https://doi.org/10.1016/j.eswa.2010.08.066>