



**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación



El futuro  
es de todos

DNP  
Departamento  
Nacional de Planeación

# Procesamiento natural del lenguaje en Planes de Desarrollo Territorial

**Unidad de Científicos de Datos**  
Dirección de Desarrollo Digital

Noviembre, 2021






- 1. Introducción**
- 2. Descripción del proyecto**
- 3. Metodología**
- 4. Resultados**
- 5. Conclusiones**



# 1. Introducción

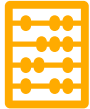
# Procesamiento natural del lenguaje en Planes de Desarrollo Territorial

<h2>¿Qué?</h2>	<p>Identificar las apuestas estratégicas registradas en los Planes de Desarrollo Territorial (PDT) 2020-2023 y la alineación de los PDT con el Plan Nacional de Desarrollo (PND) 2018-2022, mediante el uso de técnicas de análisis y minería de texto.</p>	
<h2>¿Para qué?</h2>	<p>Para facilitar el análisis de las principales temáticas abordadas en los Planes de Desarrollo Territorial por medio de gráficos y tablas de fácil interpretación.</p>	
<h2>Logros</h2>	<ol style="list-style-type: none"> <li>1. Frecuencia de términos, gráficos de barras y nubes de palabras.</li> <li>2. Búsqueda de términos de interés y articulación subnacional</li> <li>3. Términos asociados con las categorías de búsqueda</li> <li>4. Similitud PDT, estudios de caso y PND</li> <li>5. Palabras exclusivas PDT y PND</li> </ol>	

# 2. Descripción del proyecto



# Descripción del proyecto



- Identificación de las palabras y bigramas más frecuentes



- Búsqueda flexible para búsqueda de términos de interés



- Contexto de términos de interés



- Similitud entre documentos



- Palabras exclusivas

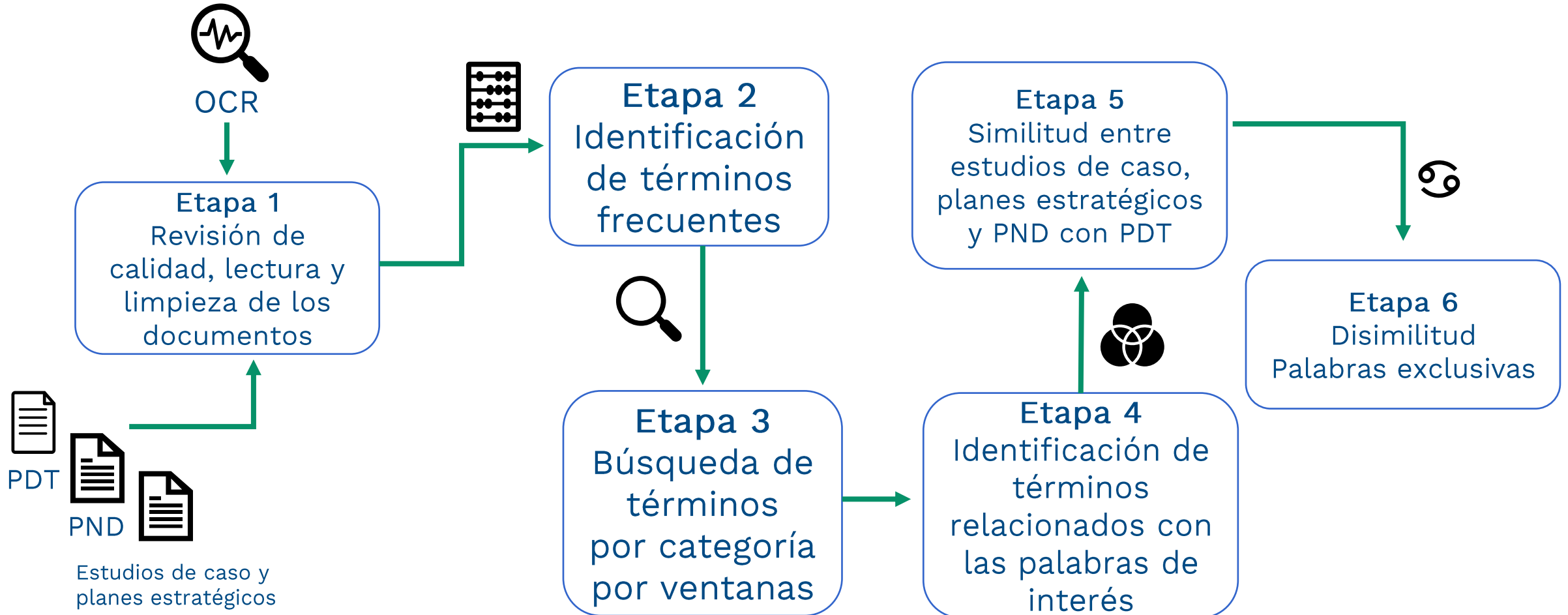


- PDT, PND, estudios de caso, planes estratégicos

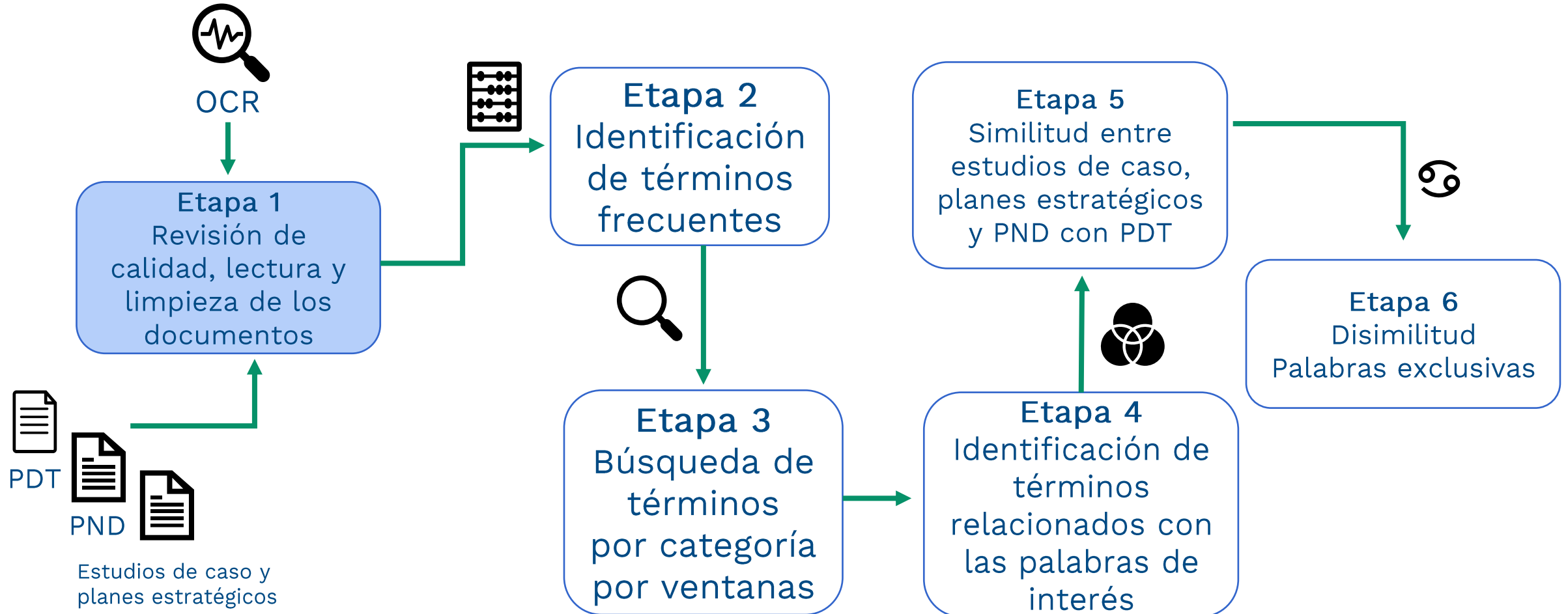
# 3. Metodología



# Metodología general



# Metodología general

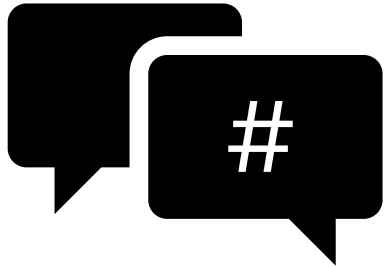


# Metodología: Lectura - Limpieza



# Metodología

## Limpieza y adecuación de textos



- 1** Remoción de números, signos de puntuación y caracteres especiales. # @ : / . ,
- 2** Eliminación de stopwords Preposiciones, artículos
- 3** Eliminación de nombres de palabras sin contexto http, com, ugihkb

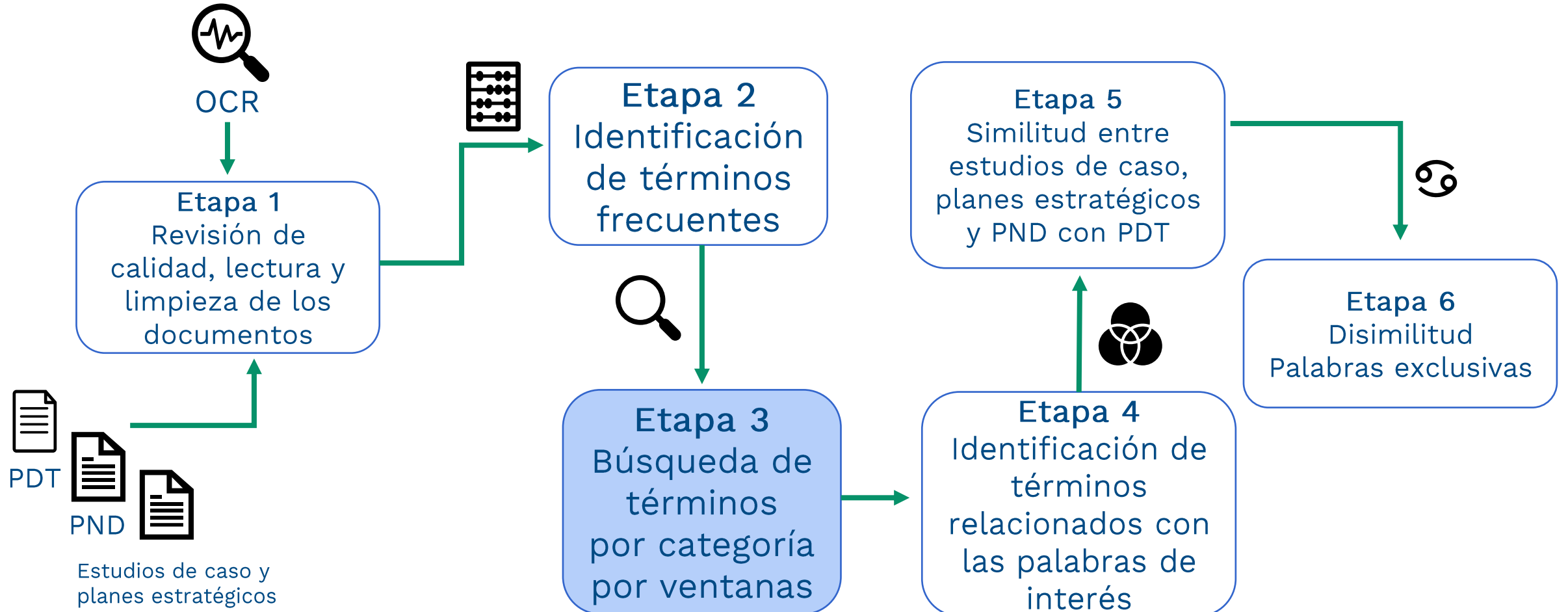
# Metodología del proyecto

## Lematización

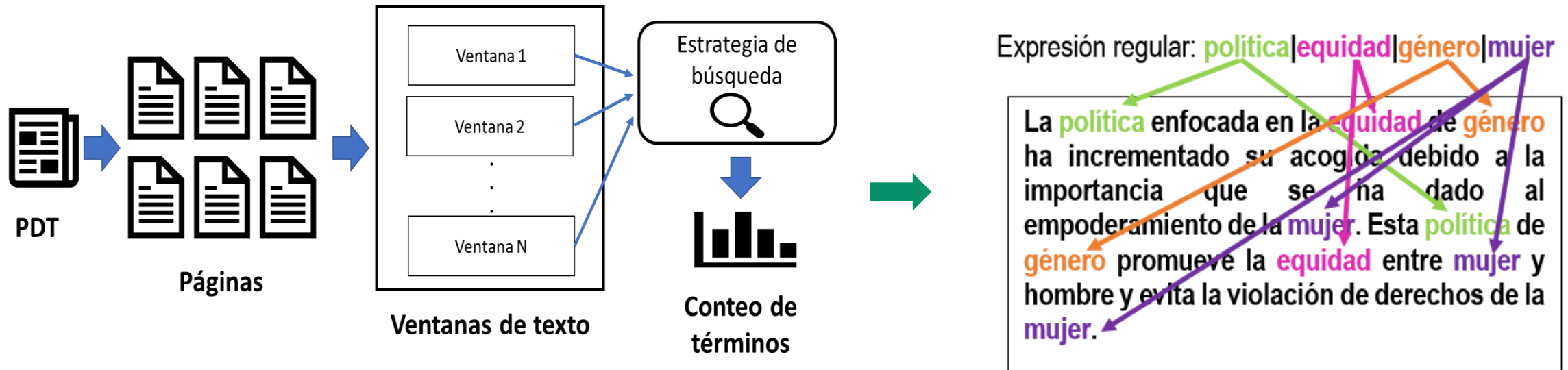


Evita tener un vocabulario más grande de lo necesario

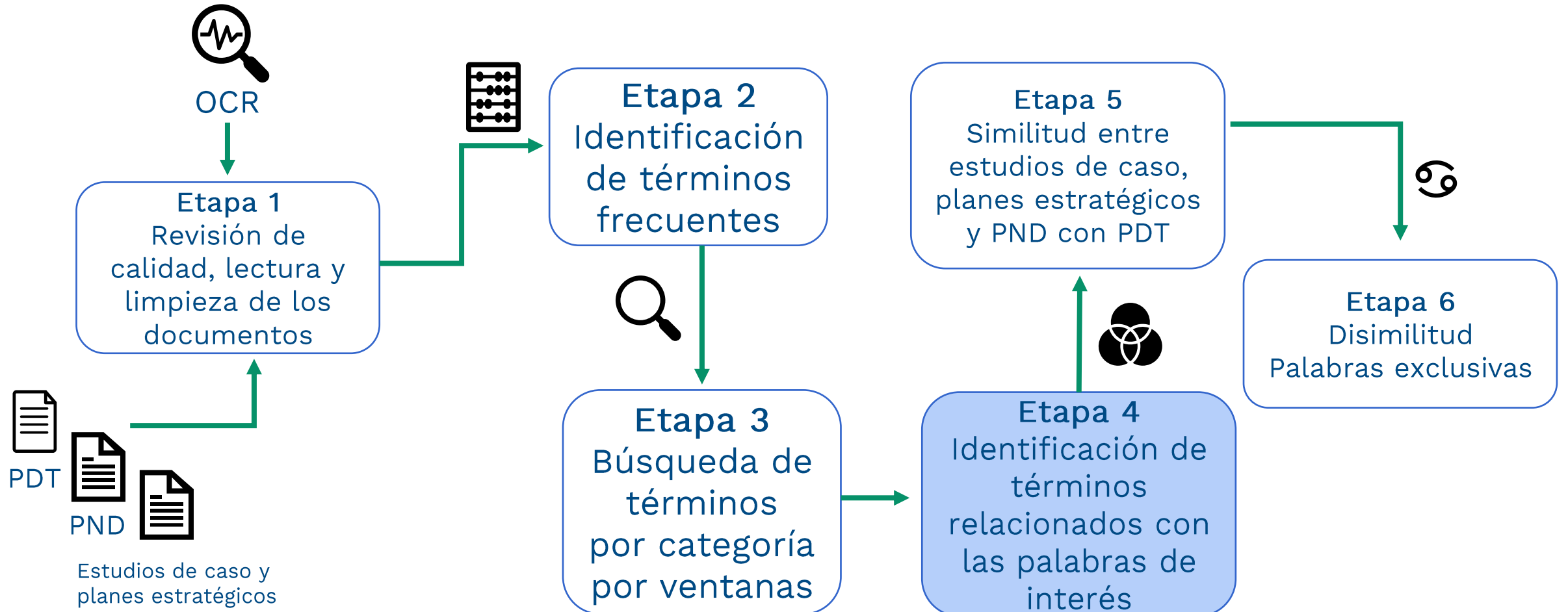
# Metodología general



# Metodología: Estrategia de búsqueda flexible



# Metodología general





# Metodología del proyecto

Búsqueda términos y términos relacionados

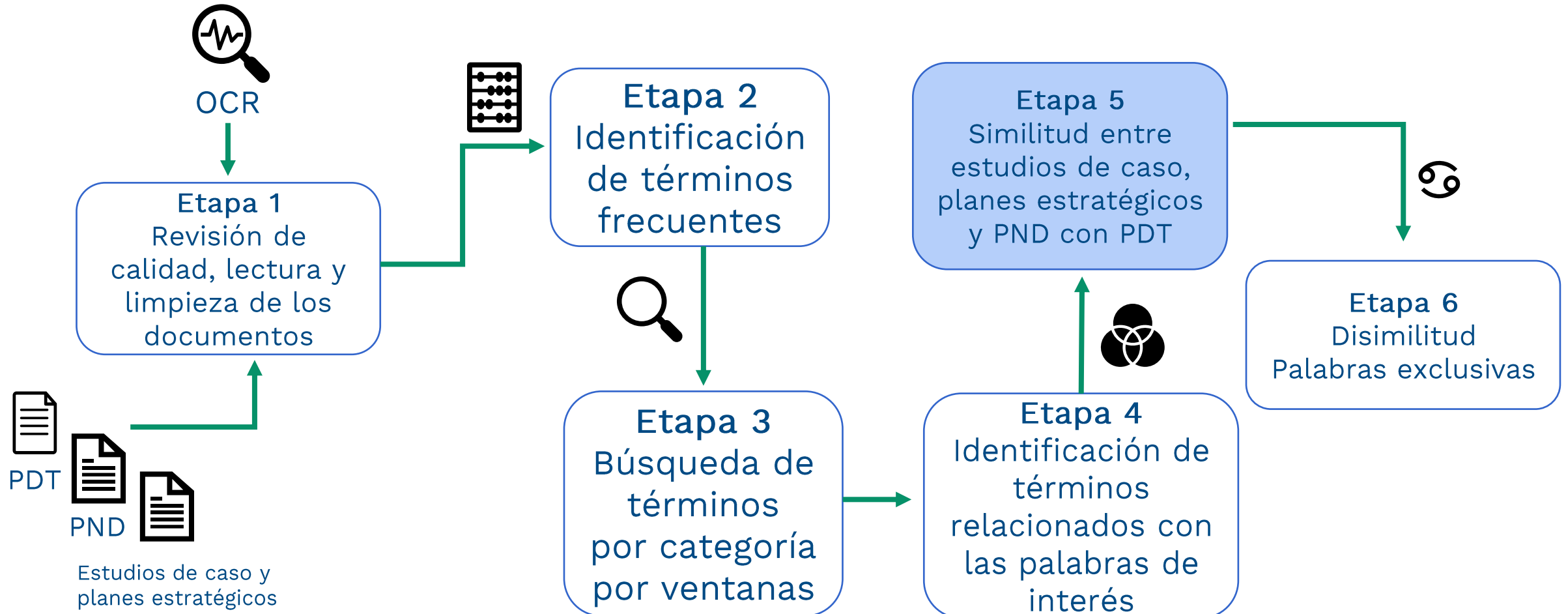


Departamento requiere disponer  
tierras aptas producción  
agropecuaria número desarrollo  
sostenibilidad condiciones vida  
adecuadas eviten migración jóvenes  
permita dirección más técnica  
procesos productivos.

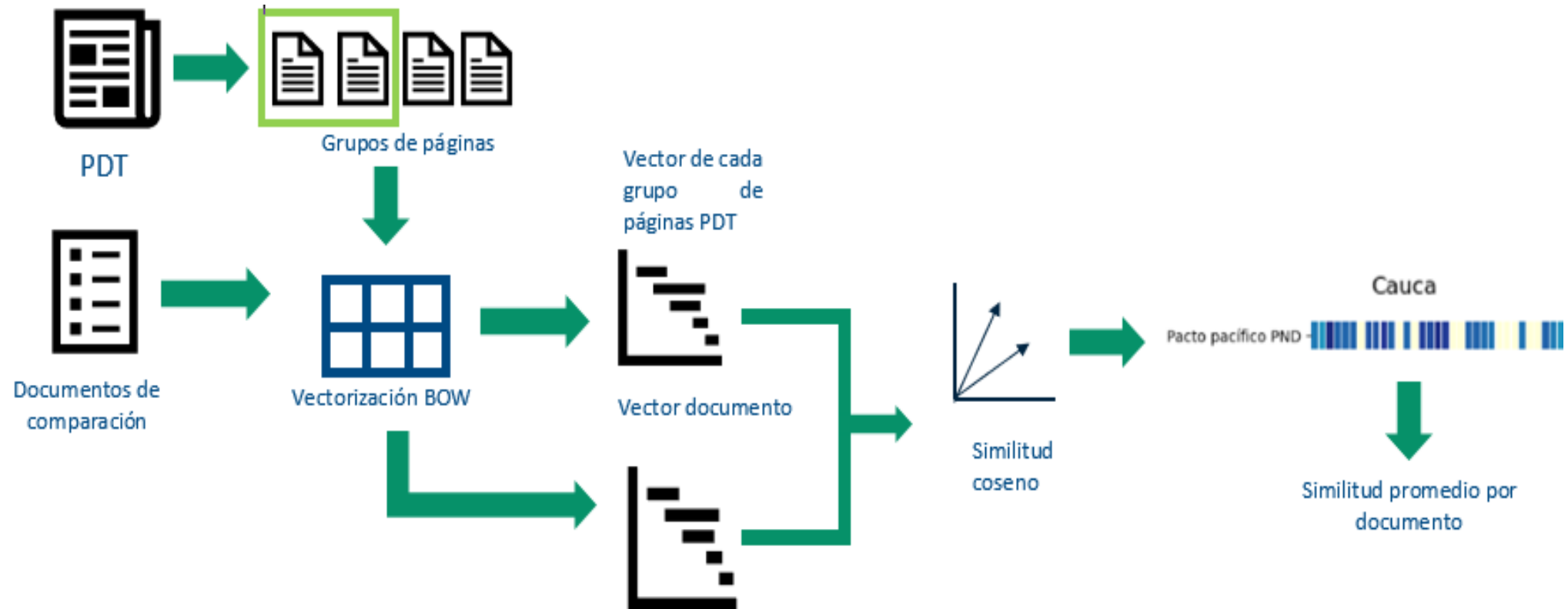
- Sostenibilidad
- Departamento
- Requiere
- Disponer
- Tierras
- .....

Agrupación por categorías

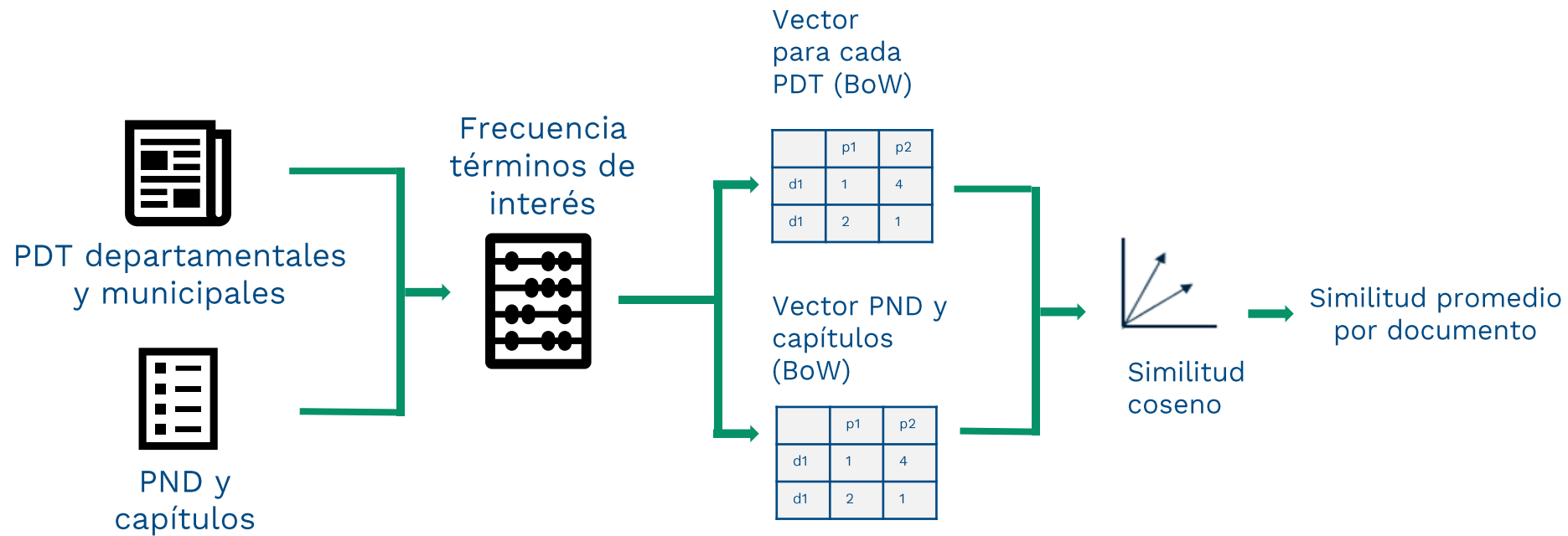
# Metodología general



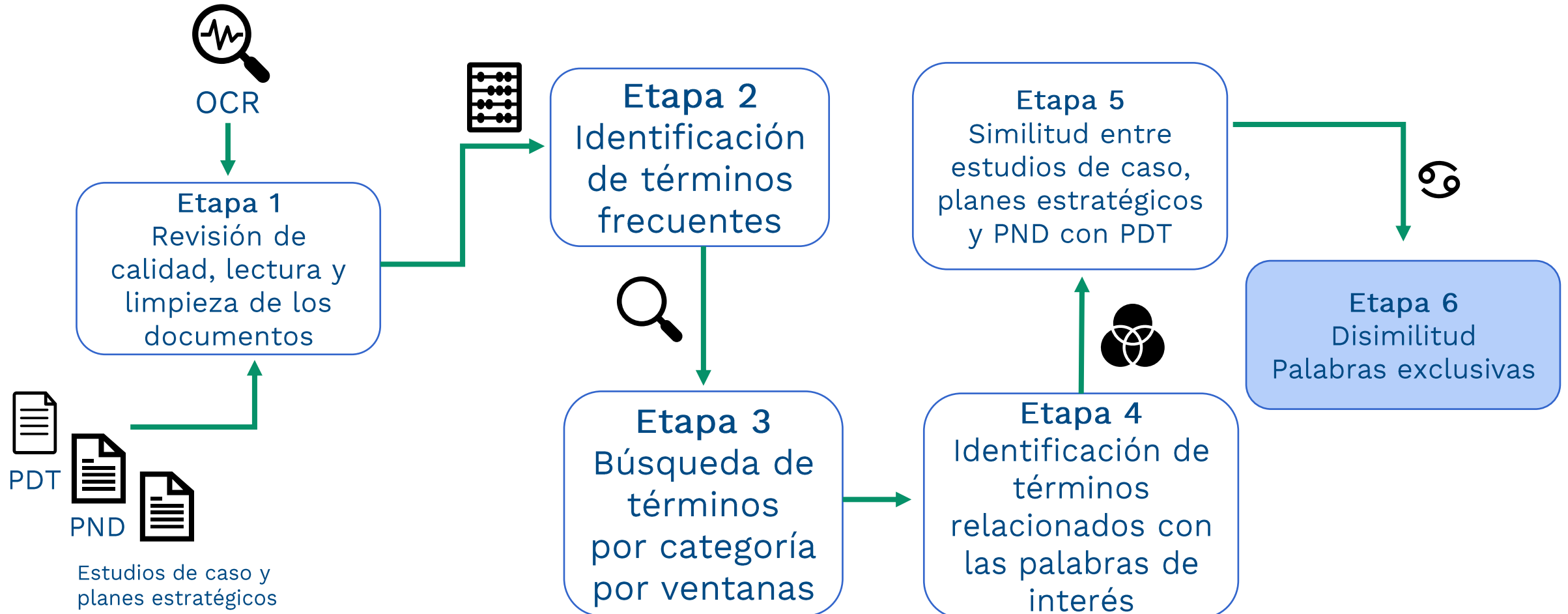
# Metodología: similitud PDT y documentos



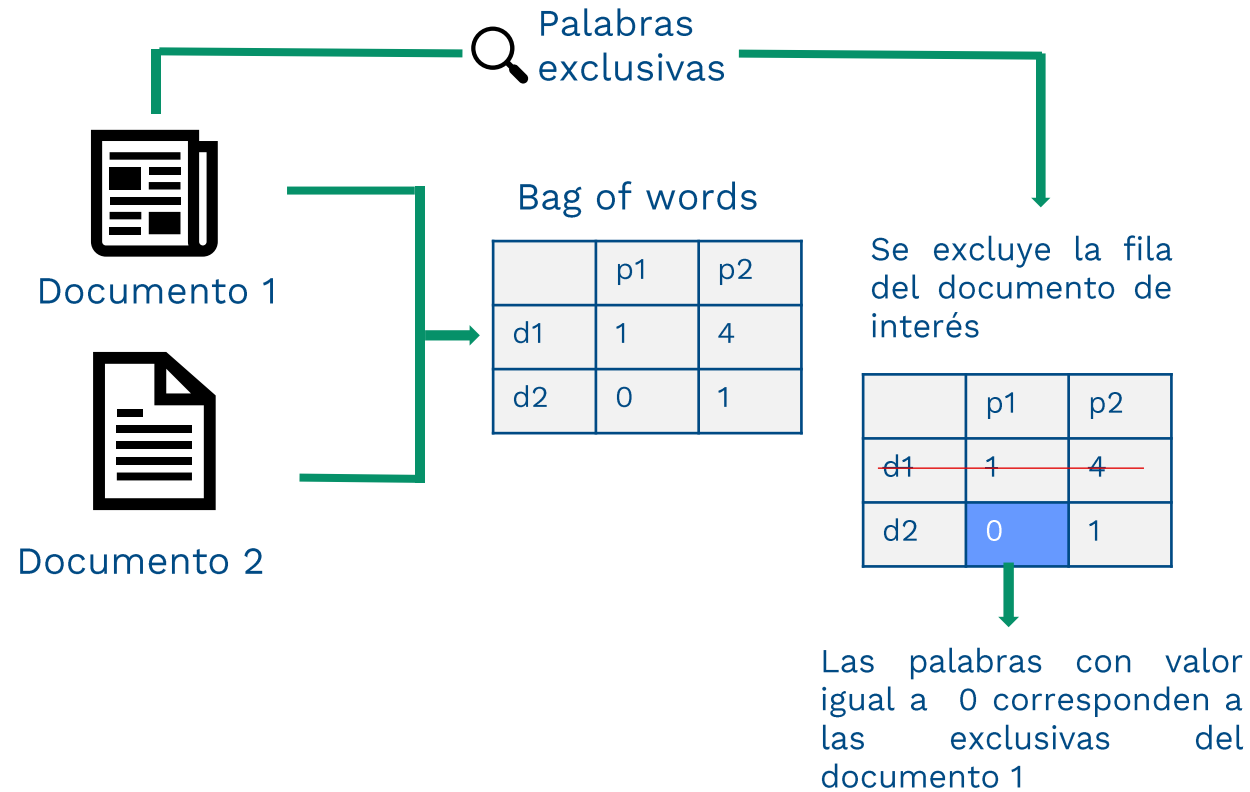
# Metodología: similitud PDT y PND



# Metodología general



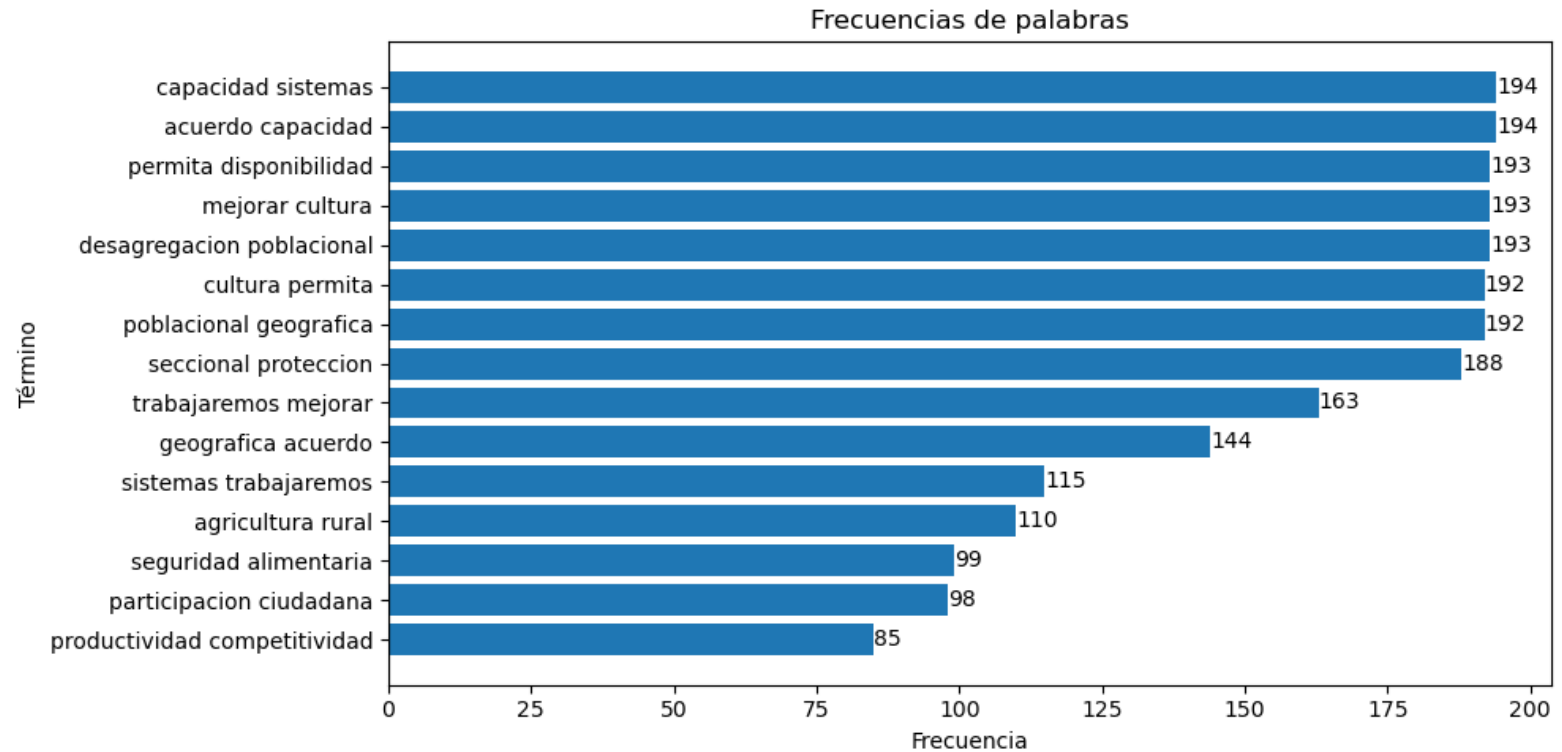
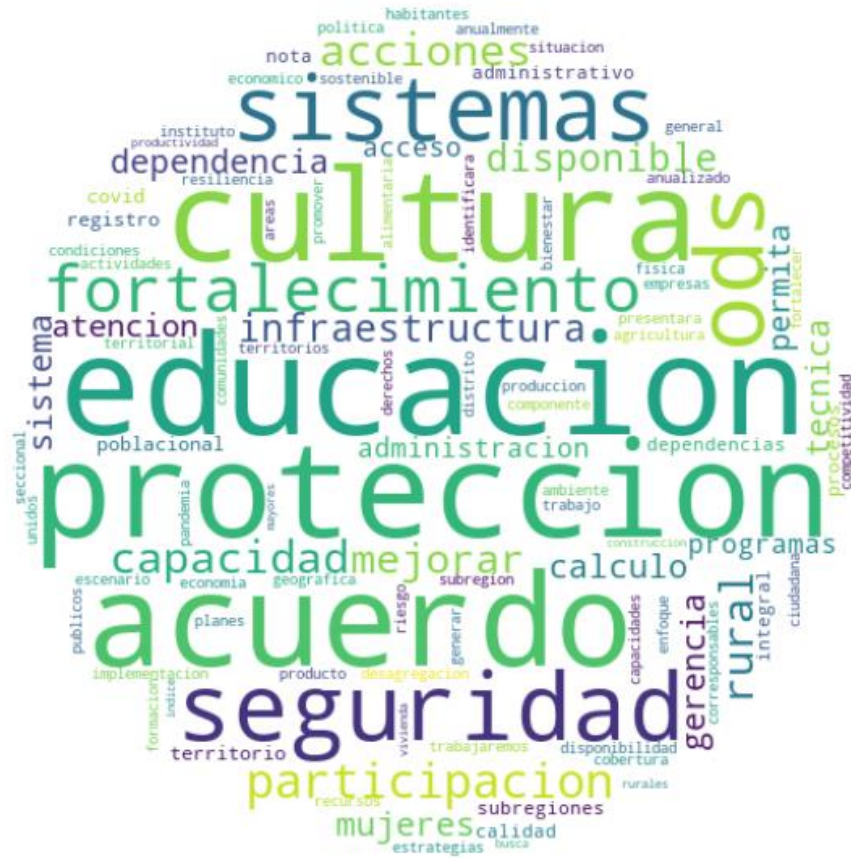
# Metodología: palabras exclusivas



# 4. Resultados

# Resultados

## Amazonas





# Resultados

## Búsqueda de términos de interés

Término	Categoría	ANTIOQUIA	ATLÁNTICO	BOLÍVAR	BOYACÁ
ADR	Agricultura y Desarrollo rural		3		4
ANT	Agricultura y Desarrollo rural	1		1	
APP	Innovación y competitividad	2	2	2	4
ARN	Paz		3		4
Acuerdo de Paris	Cambio climatico	280	44	61	77
Afiliación seguridad social	Empleo y trabajo	6		3	1
Agencia Nacional de Tierras	Agricultura y Desarrollo rural	4			
Agencia de Desarrollo Rural	Agricultura y Desarrollo rural	7	3		3

Categoría	Total
Adulto mayor	1432
Agricultura y Desarrollo rural	9579
Agua y saneamiento básico	4073
Asuntos étnicos	6145
Cambio climático	23977
Cultura	5547
Derechos	10857
Discapacidad	3069
Educación	14503
Empleo y trabajo	14455
Enfoque territorial	13837
Equidad	20727
Equidad de género	8410
Innovación y competitividad	27327
Justicia	3597
Ordenamiento territorial	6766
Participación ciudadana	4703
Paz	8166
Planeación y seguimiento	39110
Primera infancia y adolescencia	8421
Reactivación exclusivos	4720
Reactivación relevantes	23515
Recursos	7781
Salud	10104
Sector minero-energético	1812
Seguridad	5563
Transparencia y lucha contra la corrupción	7389
Transporte	918
Turismo	1008
Víctimas	891



# Resultados

Términos relacionados con las palabras de interés

Departamento	Término inicial	Categoría	Término relacionado	Cantidad
Antioquia	Sostenibilidad	Cambio climático	desarrollo	20
Antioquia	Ambiental	Cambio climático	desarrollo	20



Departamento	Categoría	Término relacionado	Cantidad
Antioquia	Cambio climático	desarrollo	40

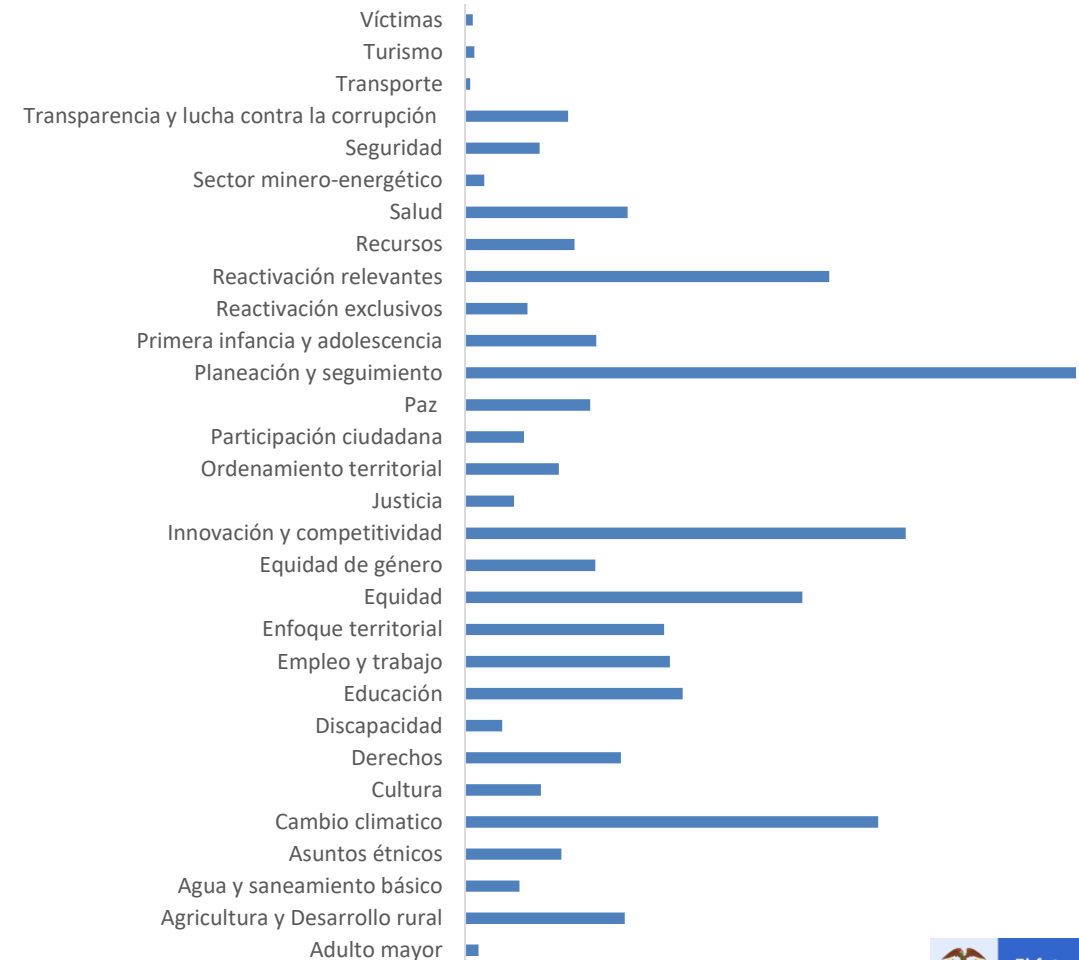


# Resultados

Por departamento

Por categoría

Departamento	ANTIOQUIA
Categoría	Adulto mayor
Palabra relacionada	Cantidad
acceso	15
acumulado	11
adulto	15
comunidad	16
concepto	21
departamento	17
desarrollo	17
financiero	15
fortalecimiento	11
funcionamiento	19
fundamental	12
información	11
ingreso	16
número	12
población	14
programa	14
situación	16
social	24
territorial	11
vulnerabilidad	11



# Resultados

## Búsqueda de términos de articulación subnacional

Término	Categoría	ANTIOQUIA	ATLÁNTICO	BOLÍVAR	BOYACÁ
Acuerdo Metropolitano	Articulación subnacional - Términos que implican articulación subnacional	5	1		
Asociatividad Territorial	Enfoque territorial - Términos que implican articulación subnacional	5			2
Conectar territorios	Enfoque territorial - Términos que implican articulación subnacional	3	5	3	7
Consejo Metropolitano	Ordenamiento territorial - Términos que implican articulación subnacional				1
Consejo Metropolitano de Planeación	Ordenamiento territorial - Términos que implican articulación subnacional				
Consejo Metropolitano de Planificación	Ordenamiento territorial - Términos que implican articulación subnacional				
EAT	Enfoque territorial - Términos que implican articulación subnacional			1	
Estrategia Metropolitana	Articulación subnacional - Términos que implican articulación subnacional	3	2		1

Categoría	Cantidad
Articulación subnacional - Términos que implican articulación subnacional	3984
Enfoque territorial - Términos que implican articulación subnacional	35463
Normatividad - Términos que implican articulación subnacional	24
Ordenamiento territorial - Términos que implican articulación subnacional	1320
Recursos - Términos que implican articulación subnacional	4518




Normalización por número de páginas, número de palabras y palabras más frecuentes



# Resultados

## Términos relacionados con las palabras de articulación subnacional

Departamento	Término inicial	Categoría	Término relacionado	Cantidad
Amazonas	Conectar territorios	Enfoque territorial - Términos que implican articulación subnacional	aati	25
Amazonas	Asociatividad territorial	Enfoque territorial - Términos que implican articulación subnacional	aati	33



Departamento	Categoría	Término relacionado	Cantidad
Antioquia	Enfoque territorial - Términos que implican articulación subnacional	aati	58

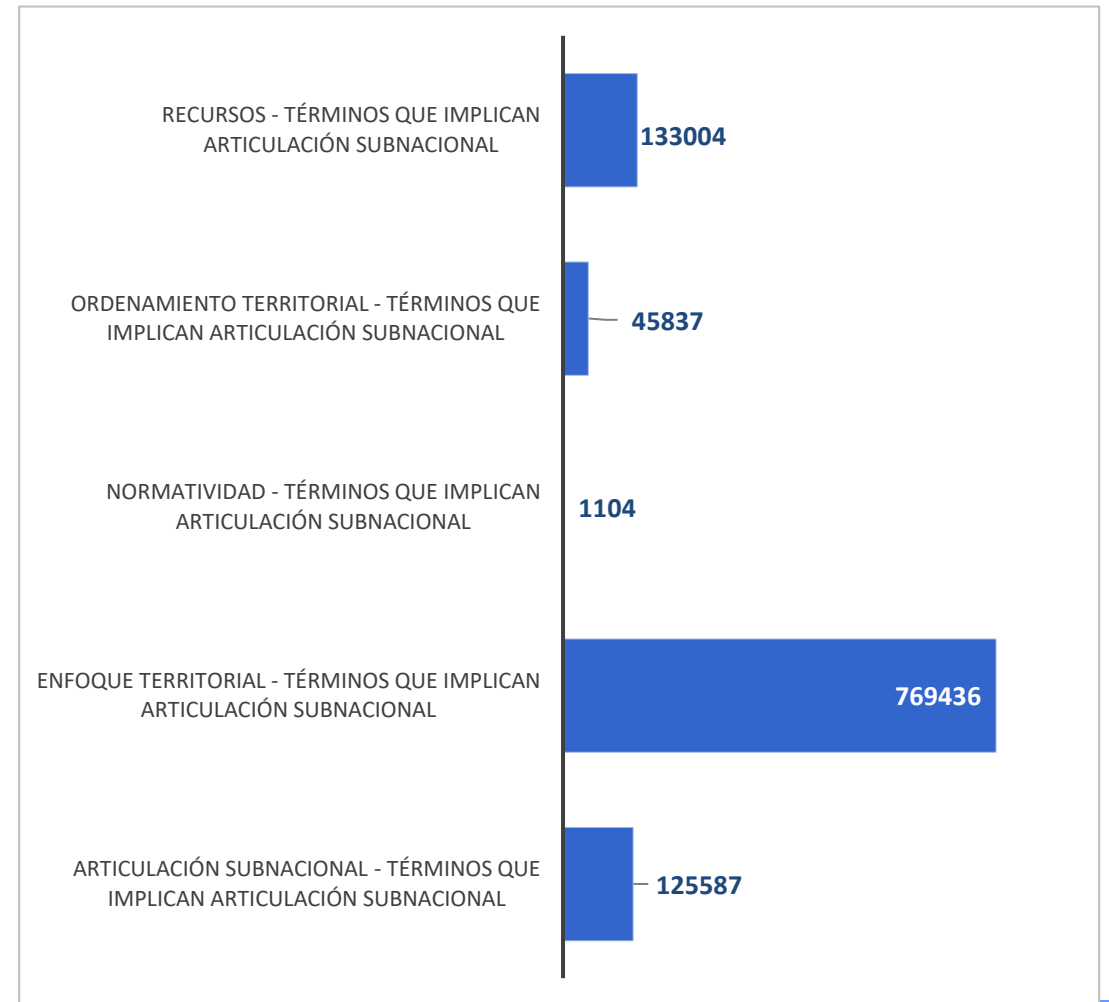


# Resultados

Por departamento

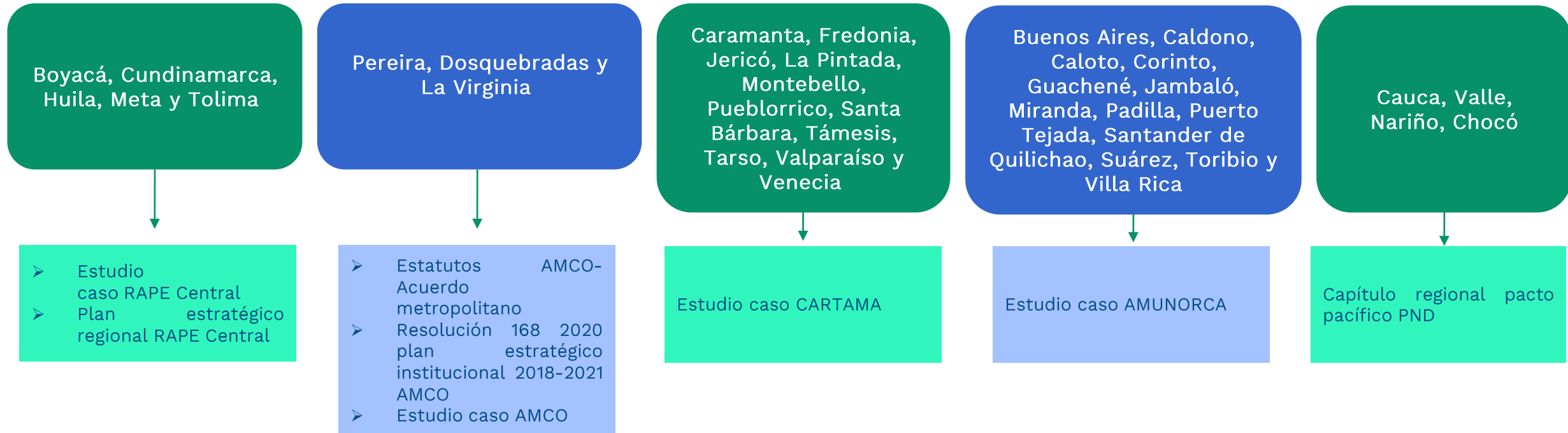
Departamento	ANTIOQUIA
Categoría	Ordenamiento territorial - Términos que implican articulación subnacional
Palabra relacionada	Cantidad
indígena	618
desarrollo	474
departamento	472
población	446
nacional	396
público	225
acción	222
gestión	211
recurso	205
contar	202
comunidad	200
social	194
nivel	193
tasa	184
educativo	183
apoyo	179
información	179
ambiental	177
gobierno	176
sistema	176

Por categoría



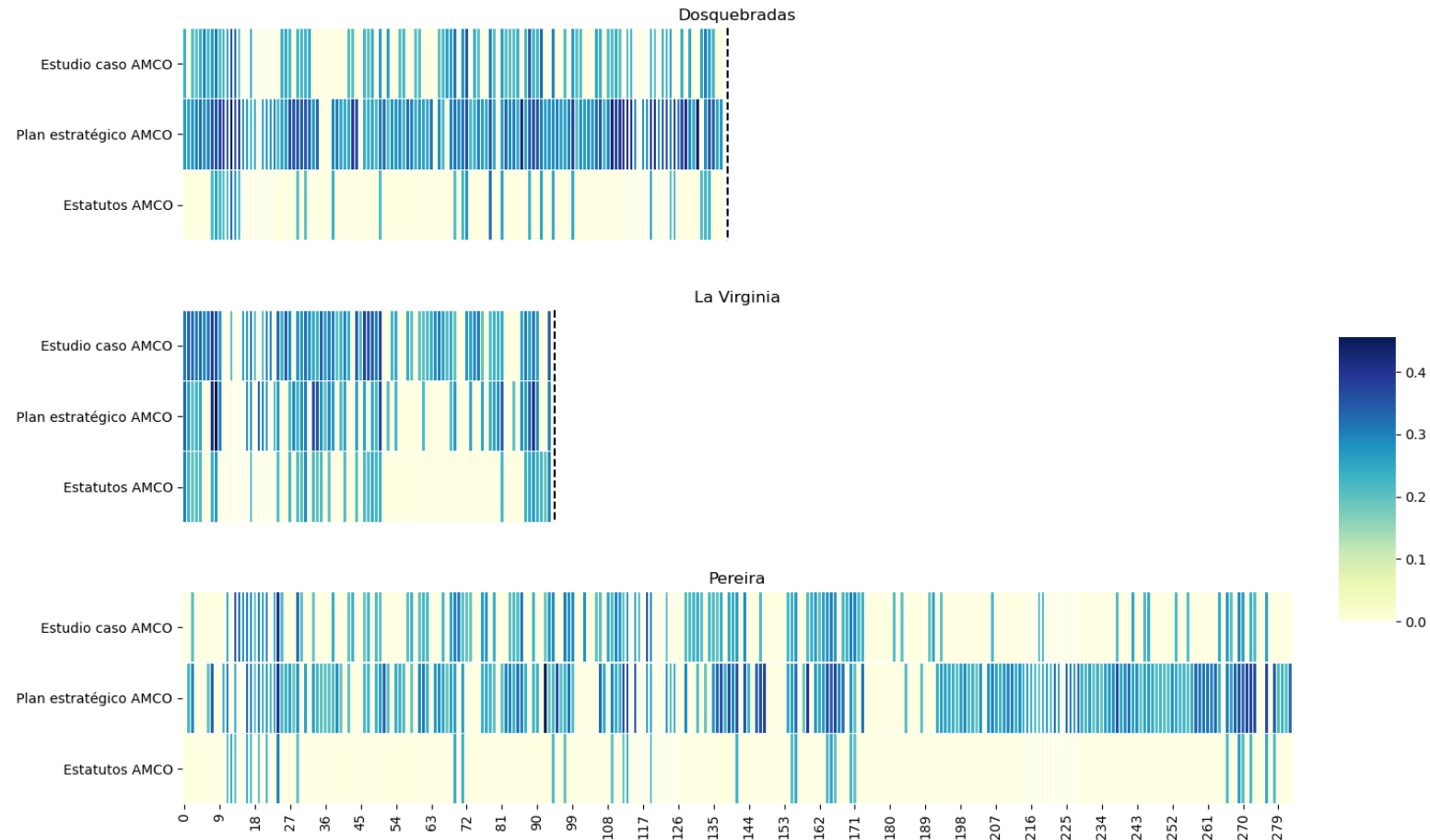
# Resultados

## Cálculo de similitud



# Resultados

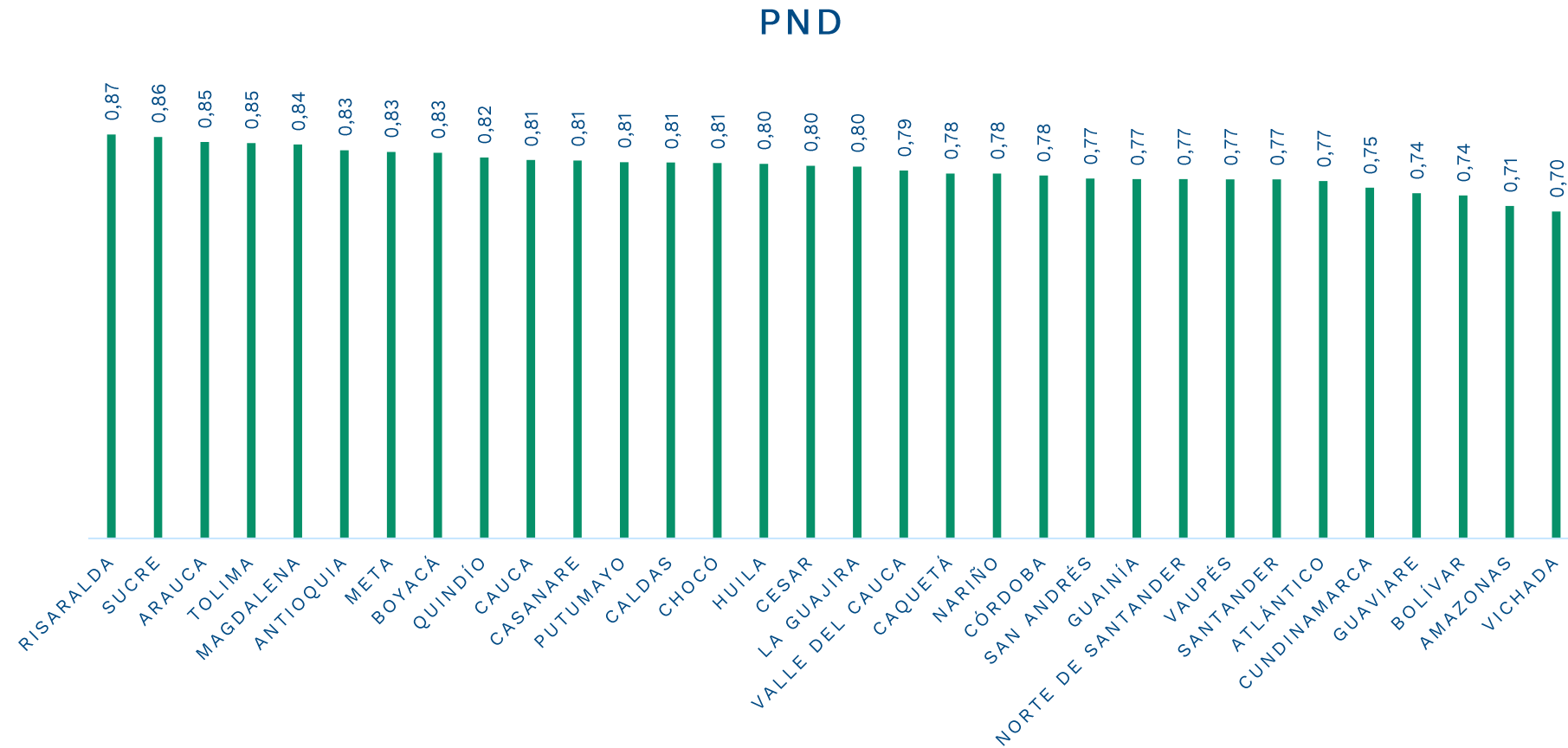
## Cálculo de similitud PDT y estudios de caso, planes estratégicos y estatutos





# Resultados

## Cálculo de similitud PDT y PND



# Resultados

Cálculo de similitud PDT y PND, promedio

PND	Promedio departamental	Promedio municipal
PND	0,79	0,69
Capítulo equidad	0,70	0,62
Capítulo legalidad	0,55	0,47
Capítulo emprendimiento	0,54	0,46
Capítulo descentralización	0,52	0,46
Capítulo productividad	0,37	0,31
Capítulo construcción paz	0,52	0,46
Capítulo equidad mujeres	0,39	0,33
Capítulo equidad grupos	0,44	0,38



# Resultados

Cálculo de similitud PDT y PND, PDT más similares

PND	PDT	Similitud	PDT	Similitud
PND	Risaralda	0,87	Génova	0,87
Capítulo equidad	Cesar	0,78	Montería	0,83
Capítulo legalidad	Risaralda	0,65	Copacabana	0,70
Capítulo emprendimiento	Meta	0,66	Dosquebradas	0,65
Capítulo descentralización	Quindío	0,69	Sopó	0,72
Capítulo productividad	Guainía	0,49	El Rosario	0,59
Capítulo construcción paz	Arauca	0,62	Roberto Payán	0,64
Capítulo equidad mujeres	Chocó	0,56	Casabianca	0,71
Capítulo equidad grupos	Chocó	0,55	Guapi	0,59



# Resultados

## Palabras exclusivas PND

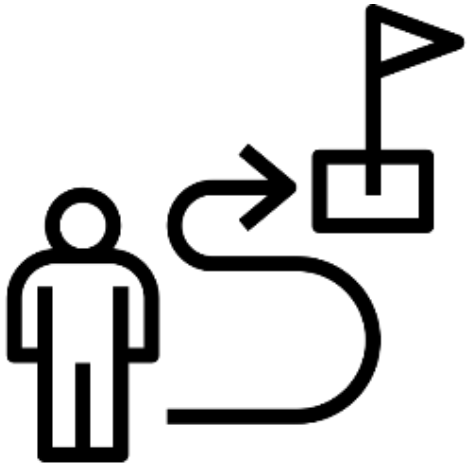
Capítulo equidad	Capítulo legalidad	Capítulo emprendimiento	Capítulo descentralización	Capítulo productividad	Capítulo construcción paz	Capítulo equidad mujeres	Capítulo equidad grupos
vejez	consulado	mincomercio	inmueble	malformación	guardar	empareja	tentación
saludable	lavado	fitosanitario	macroproyecto	pirámide	antipersona	ingeniería	afrocolombiano
prejuicio	blando	mercantil	equilibrado	protesta		igualitario	indígenas
deportista	judicialización	intangible	suburbanización	traza		reproducción	indigno
barrio	terrorismo	bancario	aeródromo			anticonceptivo	nativo
rendimiento	anticorrupción	firmado	edificabilidad			marital	amazónica
recreativo	carcelario	patente	ecosistémico			feminidad	restauración
federación	dni	procolombia	aeropuerto			casada	sagrado
malnutrición	disfunción	sanidad	obsolescencia			sobrecarga	tradición
propietario	ciberespacio	microfinanza	multipropósito			popular	bélica



# 5. Conclusiones y recomendaciones



# Conclusiones



1

Se desarrolló una metodología de búsqueda de términos relacionados con cada tema basado en expresiones regulares, lo que permite encontrar múltiples formas de una frase en los textos.

2

El PDT departamental que mayor similitud presenta con el PND respecto al uso de los términos de las 30 categorías es el de Risaralda, en cuanto al municipal es el PDT de Génova.

3

El promedio de similitud de todos los PDT departamentales con respecto al PND es 0,79 y a nivel municipal 0,69.

4

El capítulo de productividad del PND es el que presenta menor similitud en promedio con respecto a los PDT departamentales y municipales, por otra parte el que mayor similitud presenta es el capítulo de equidad.

# Conclusiones

5

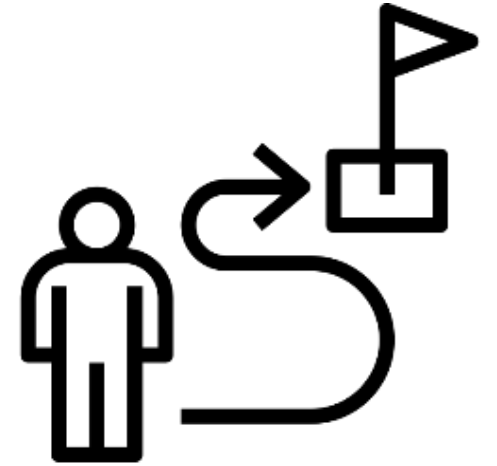
Las palabras exclusivas son una herramienta para diferenciar el contenido de los diferentes documentos analizados.

6

La lectura y limpieza de los documentos no arroja resultados perfectos, esto ocasiona que los textos de análisis contengan palabras que no tienen contexto, con el fin de disminuir estas palabras se llevó a cabo una corrección ortográfica que incrementó el tiempo de desarrollo.

7

Se recomienda validar los resultados de las palabras exclusivas y de todos los resultados desde la vista de un experto dado que las metodologías de lectura y limpieza de los textos pueden ocasionar errores.





**El futuro  
es de todos**

**DNP**  
Departamento  
Nacional de Planeación