



GOBIERNO  
DE COLOMBIA



DNP DEPARTAMENTO  
NACIONAL  
DE PLANEACIÓN

Departamento Nacional de Planeación

[www.dnp.gov.co](http://www.dnp.gov.co)

# SIMILITUD DE PROYECTOS TIPO

Unidad de Científicos de Datos  
Dirección de Desarrollo Digital

 @DNP\_Colombia

Noviembre, 2018  
dnp.gov.co



**DNP** DEPARTAMENTO  
NACIONAL  
DE PLANEACIÓN

1. Motivación
2. Metodología
3. Resultados
4. Conclusiones



# SIMILITUD DE PROYECTOS TIPO

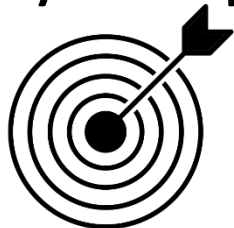
# Motivación

1

## SIMILITUD DE PROYECTOS TIPO

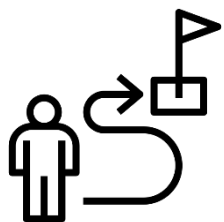
# Algoritmo de similitud entre campos de texto

Herramienta que permite detectar el número de proyectos nuevos que siguen los lineamientos de los Proyectos Tipo, establecidos en el documento CONPES 3856 de 2016



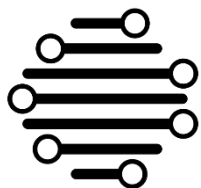
## Objetivo

Estimar una medida similitud entre campos de texto de proyectos nuevos y Proyectos Tipo financiados a través del Sistema General de Regalías - SGR



## Metodología

Similitud de proyectos de diferentes sectores con respecto a proyectos estandarizados basado en la vectorización de palabras para cada campo de texto



## Insumos

**Texto extraído de 4047 proyectos nuevos** – 8953 palabras en promedio por sector

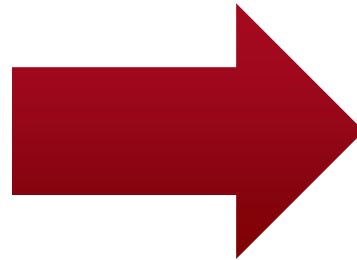
**20 Proyectos Tipo de 7 sectores** cuya Metodología General Ajustada - MGA se encuentra disponible en la página web del DNP

# Definición de Proyecto Tipo

Es una **estandarización** de los proyectos del Sistema General de Regalías – SGR del que pueden hacer uso las entidades territoriales

## Problema

**Confusión y desinformación** de las entidades que genera desigualdad de estándares



## Solución

Las entidades ahora pueden atender problemáticas específicas de manera **ágil y eficiente**

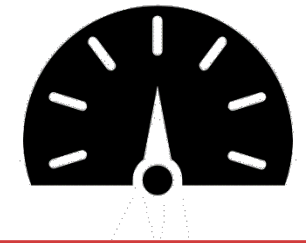
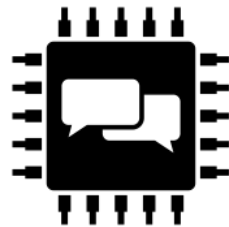
# Metodología

2

## SIMILITUD DE PROYECTOS TIPO

# Diagrama de trabajo para el análisis de los proyectos

Se realiza una selección de los proyectos que pertenezcan a los 7 sectores para los cuales existen Proyectos Tipo. Posteriormente se lleva a cabo una depuración del texto y finalmente, se calcula el **índice de similitud** entre proyectos nuevos y Proyectos Tipo



Filtro por sector

Preprocesamiento del  
texto

Análisis de similitud

Algoritmo de  
resultados finales



# Definición de Proyectos Tipo para el contraste

Un Proyecto Tipo tiene las siguientes características que son utilizadas para construir el algoritmo de similitud

## 11 campos de texto en común

- Nombre
- Estado\*
- Sector\*
- Descripción Problema Necesidades
- Descripción Necesidades
- Objetivo
- Análisis Técnico
- Nombre
- Problema Central
- Producto
- Actividad

## 7 sectores seleccionados



Cultura (2)



Transporte (7)



Ciencia I+D (2)



Educación (2)



Salud y protección social (2)



Deporte y recreación (3)



Agricultura y desarrollo rural (2)

\*Estos dos campos de texto se descartan puesto que no brindan información útil para el problema

# Proyectos Tipo para contrastar

Se encuentra un total de 20 Proyectos Tipo distribuidos en los 7 sectores seleccionados y son los siguientes



Transporte (7)

- Pavimento rígido
- Cicloinfraestructura
- Mejoramiento vías
- Mejoramiento vías y caracterización
- Inventario vial
- Puentes
- Placa huella



Educación (2)

- Transporte escolar
- Tabletas



Ciencia  
I+D (2)

- Váuchers de innovación
- Ondas



Agricultura  
y desarrollo rural (2)

- Construcción viviendas
- Unidades sanitarias



Cultura (2)

- Biblioteca
- Escuela de música



Salud  
y protección social (2)

- Cadena de frío
- Dotación ambulancias



Deporte  
y recreación (3)

- Parque recreodeportivo
- Placa deportiva
- Cancha en grama sintética

# Proceso de filtrado y selección de términos

Se realiza la construcción de un algoritmo para determinar las palabras adecuadas de **comparación para cada campo y sector de la muestra**



Proceso iterativo para **cada campo de texto** en cada uno de los sectores



# Ejemplo del funcionamiento del algoritmo

Para cada campo de texto se encuentra un **porcentaje de similitud** utilizando la lista de términos creada previamente bajo la metodología de comparación formulada

## Proyecto nuevo para el sector transporte

**Nombre:** **Servicio** de **transporte escolar** para los estudiantes en la zona rural con limitaciones de transporte para acceder a las diferentes sedes educativas vigencia 2018 en el Municipio de Yotoco

## Proyecto Tipo 1 para el sector transporte

**Nombre:** Prestación del **servicio** de **transporte escolar** en Meta

## Fórmula de comparación

$$s_{li} = 1 - \frac{\left| \sum_{j=1}^J (x_{ij} - y_{lj}) \right|}{\sum_{j=1}^J (x_{ij} + y_{lj})}$$



Comparación con la lista de **términos** generados previamente

0% indica que los dos proyectos son disímiles



La similitud para el campo nombre es de un **42%**

# Resultados

3

## SIMILITUD DE PROYECTOS TIPO

# Resultado con un umbral de similitud del 70%

De 2972 proyectos analizados, el 3.5% (105 proyectos) supera el umbral estipulado

Sector	Total de proyectos	Tipo 1	Tipo 2	Tipo 3	Tipo 4	Tipo 5	Tipo 6	Tipo 7	Resultado
Transporte	1523	8	0	17	3	4	2	53	87
Ciencia I+D	122	3	2	-	-	-	-	-	5
Educación	408	7	0	-	-	-	-	-	7
Agricultura y Desarrollo rural	271	5	2	-	-	-	-	-	7
Cultura	109	1	1	-	-	-	-	-	2
Salud y Protección social	151	0	2	-	-	-	-	-	2
Deporte y Recreación social	388	0	8	5	-	-	-	-	13
<b>Total de proyectos</b>	<b>2972</b>	-	-	-	-	-	-	-	<b>123</b>

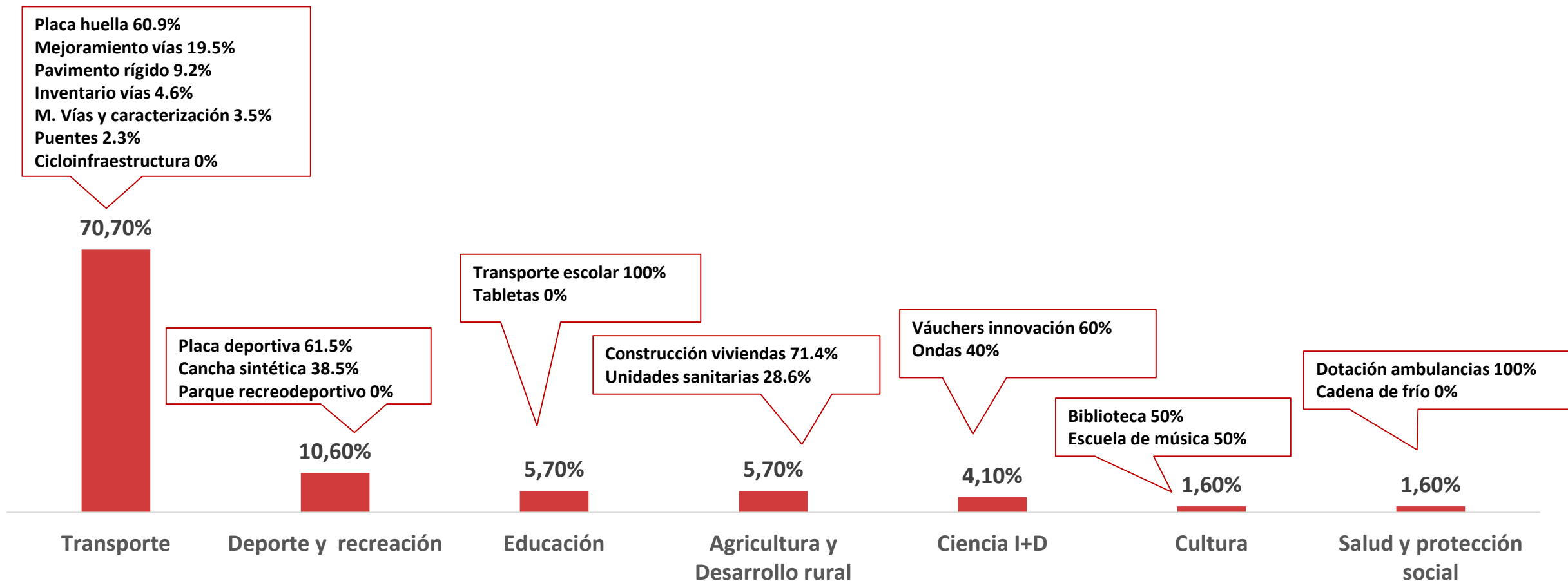
Se presentan 53 documentos que superan el umbral propuesto para el tipo 7 de transporte, que corresponde al Proyecto Tipo: Placa huella

Se presentan 123 coincidencias para un total de **105 proyectos**, puesto que un proyecto nuevo puede superar el umbral de hasta 3 Proyectos Tipo

- Significa la ausencia de proyectos para este tipo

# Porcentaje de participación por sector con umbral del 70%

El sector Transporte representa el 70.7% de las coincidencias encontradas para el umbral del 70% de similitud. Siendo Placa huella el Proyecto Tipo con mayor participación de este sector (60.9%)



# Resultado con un umbral de similitud del 90%

De 2972 proyectos analizados, el 1% (29 proyectos) supera el umbral estipulado

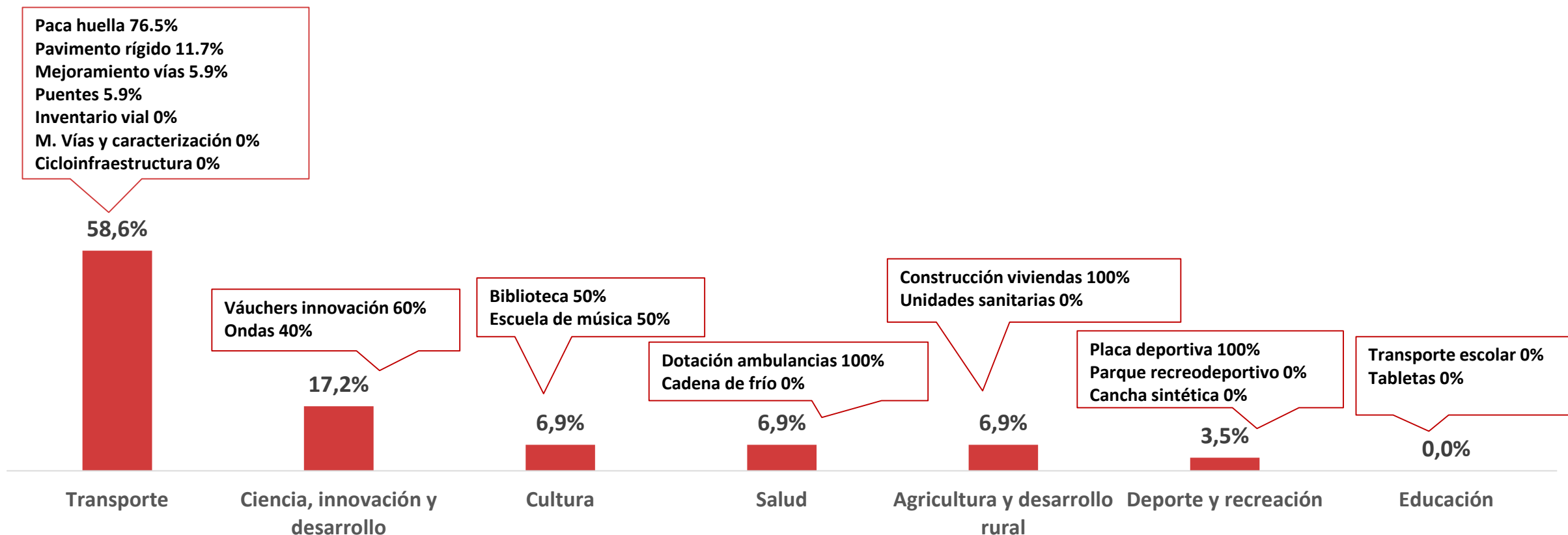
Sector	Total de proyectos	Tipo 1	Tipo 2	Tipo 3	Tipo 4	Tipo 5	Tipo 6	Tipo 7	Resultado
Transporte	1523	2	0	1	0	0	1	13	17
Ciencia I+D	122	3	2	-	-	-	-	-	5
Educación	408	0	0	-	-	-	-	-	0
Agricultura y Desarrollo rural	271	2	0	-	-	-	-	-	2
Cultura	109	1	1	-	-	-	-	-	2
Salud y Protección social	151	0	2	-	-	-	-	-	2
Deporte y Recreación social	388	0	1	0	-	-	-	-	1
<b>Total de proyectos</b>	<b>2972</b>	-	-	-	-	-	-	-	<b>29</b>

- Significa la ausencia de proyectos para este tipo



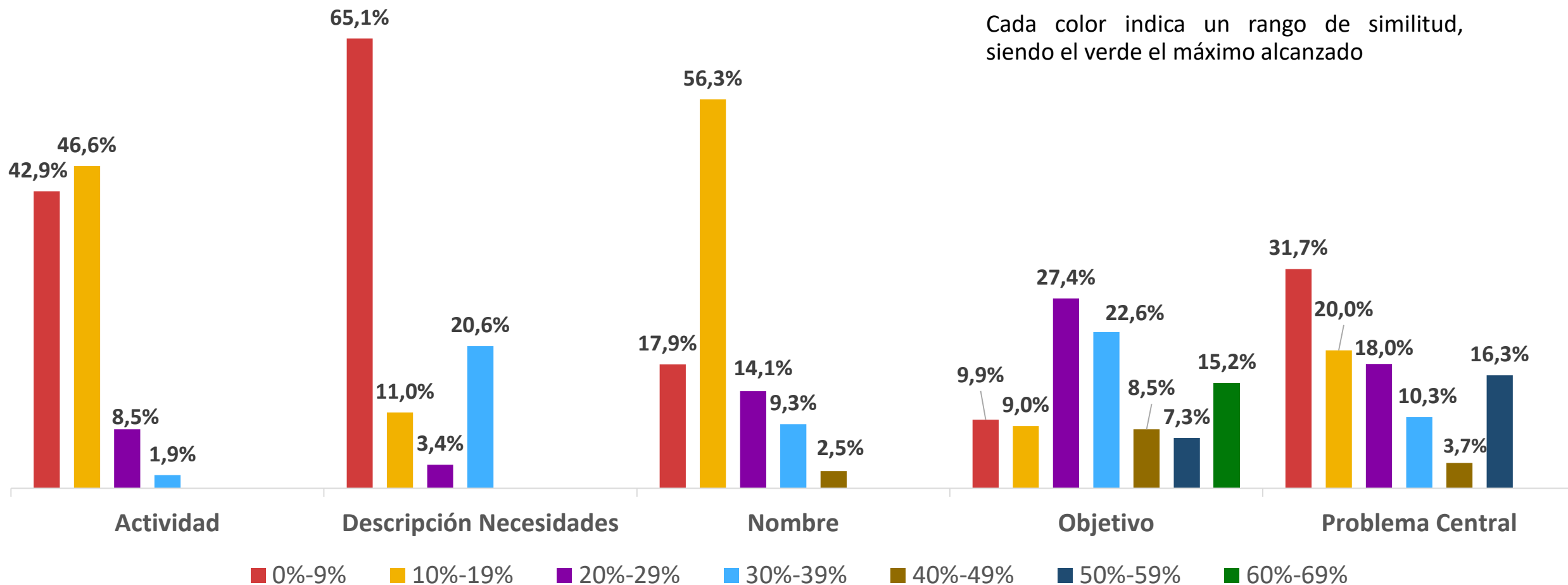
# Porcentaje de participación por sector con umbral del 90%

El sector Transporte representa el 58.6% de las coincidencias encontradas para el umbral del 90% de similitud. Siendo Placa Huella el Proyecto Tipo con participación de este sector (76.5%)



# Similitud total para el sector transporte

El siguiente gráfico muestra 5 campos del **sector Transporte** que posee el **mayor número de Proyectos Tipo**, donde hay 1523 de estos. Se observa que el **15.2%** de proyectos nuevos del campo Objetivo supera el 60% de similitud



# Conclusiones

4

## SIMILITUD DE PROYECTOS TIPO

# Conclusiones

- Tras la consulta de múltiples técnicas de cálculo de similitud entre campos de texto, se optó por realizar una comparación palabra a palabra que descartara la aparición en la mayor medida posible de falsos positivos
- Un bajo porcentaje de las entidades territoriales están siguiendo los lineamientos para los cuales los Proyectos Tipo fueron creados
- Es probable que la falta de información con respecto al uso de los Proyectos Tipo sea la razón por la que el porcentaje de similitud sea tan bajo

# Aplicaciones

- Detección de patrones en el uso adecuado de la Metodología General Ajustada para proyectos estandarizados del Sistema General de Regalías
- Control y seguimiento de los objetivos establecidos por el documento CONPES 3856 de 2016
- Cálculo de similitud entre documentos de política pública que deberían seguir el formato de un documento estándar